

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

(19)



Europäisches Patentamt

Eur pean Patent Office

Office eur péen des brevets



(11)

EP 0 781 848 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:

02.07.1997 Bulletin 1997/27

(21) Application number: 96120899.8

(22) Date of filing: 27.12.1996

(51) Int. Cl.<sup>6</sup>: C12N 15/62, C12N 9/12,  
C12N 15/48, C07K 14/195,  
C07K 14/20, C12N 15/54,  
C07K 14/15, C12N 15/31

(84) Designated Contracting States:

AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE

(30) Priority: 28.12.1995 JP 352225/95

(71) Applicant: FUJIREBIO Inc.

Shinjuku-ku, Tokyo 163-07 (JP)

(72) Inventors:

• Ueno, Eichi,  
c/o Fujirebio Inc.  
Shinjuku-ku, Tokyo (JP)

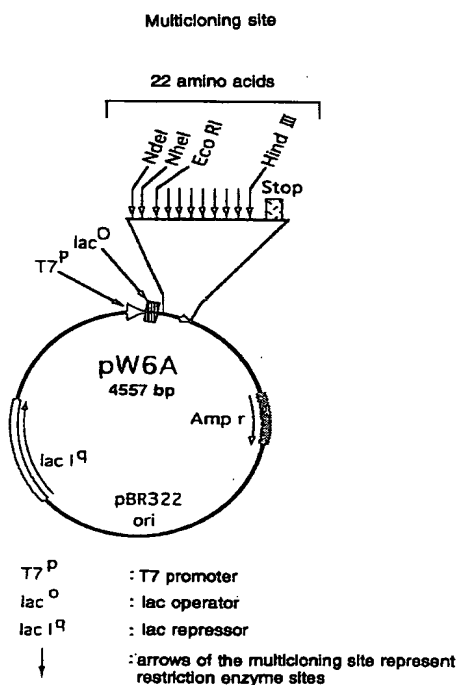
• Fujii, Nobuyuki,  
c/o Fujirebio Inc.  
Shinjuku-ku, Tokyo (JP)  
• Okada, Masahisa,  
c/o Fujirebio Inc.  
Shinjuku-ku, Tokyo (JP)

(74) Representative: Hansen, Bernd, Dr. Dipl.-Chem.  
et al  
Hoffmann, Eitle & Partner,  
Patentanwälte,  
Arabellastrasse 4  
81925 München (DE)

(54) Fused DNA sequence, fused protein expressed from said fused DNA sequence and method for expressing said fused protein

(57) Disclosed are a fused DNA sequence which comprises a DNA sequence of a heat-resistant protein, fused directly or indirectly to a DNA sequence coding a selected protein or peptide, a fused protein expressed from the fused DNA sequence, and a method for expressing the fused protein.

Fig. 1



## Description

## BACKGROUND OF THE INVENTION

5 This invention relates to expression of a fused protein, more specifically to a fused DNA sequence including a DNA sequence coding a heat-resistant protein, a fused protein expressed by said fused DNA sequence, and a method for expressing said fused protein.

Progress in genetic engineering has enabled analysis of a protein which has been purified from a natural substance, at a genetic level and artificial amplification of a desired protein (Itakura et al., Science, vol. 198, p. 1056 (1977)). By application of a DNA sequence to which thioredoxin (hereinafter referred to as "TRX" in the specification) (International Provisional Patent Publication No. 507209/1993) or glutathione-S-transferase (hereinafter referred to as "GST" in the specification) (International Provisional Patent Publication No. 503441/1989) which has been invented thereafter is fused, even a protein which is inherently expressed with difficulty can be expressed, and a technique of expressing a fused protein has been used widely.

15 TRX and GST can be applied to fusion and expression of various proteins which are expressed with difficulty, but even in GST which has been essentially used for the purpose of expressing a soluble fused protein, a fused protein becomes insoluble depending on a protein to be fused so that productivity is lowered, or a fused protein to which TRX is fused may have a problem that a nonspecific reaction is liable to occur. Therefore, it has been desired to provide a fused protein having further excellent operatability and productivity.

## SUMMARY OF THE INVENTION

Thus, an object of the present invention is to provide a novel fused DNA sequence having excellent operatability and productivity for expressing a desired protein or peptide, a fused protein expressed from said fused DNA sequence, and a method for expressing the fused protein using said fused DNA sequence.

25 The present inventors have studied intensively in order to solve the problems in the art and consequently found that when a DNA sequence coding a selected protein or peptide and a DNA sequence coding a heat-resistant protein are fused directly or indirectly and a fused protein is expressed from the resulting fused DNA sequence, the productivity of the desired protein or peptide is raised, and said fused protein has heat resistance to make a purification step simple and easy, to accomplish the present invention.

30 That is, the present invention relates to a fused DNA sequence comprising a DNA sequence coding a heat-resistant protein or peptide, fused directly or indirectly to a DNA sequence coding a selected protein or peptide, a fused protein expressed by said fused DNA sequence, and a method for expressing the fused protein using said DNA sequence.

35 The fused protein of the present invention has high solubility and can maintain even heat resistance derived from heat-resistant protein genes. Because of such a characteristic of the fused protein, when the fused protein is purified, unnecessary substances can be removed simply and easily by heat treatment so that the fused protein can be obtained with good yield.

40 In the case of TRX derived from *Escherichia coli* and GST derived from *Schistosoma japonicum*, which have been widely used as a fused protein, *Escherichia coli* and *Schistosoma japonicum* can live in bodies of mammals and other creatures so that when a fused protein using TRX or GST is used as an antigen of an immunoreaction, a nonspecific reaction due to *Escherichia coli* or *Schistosoma japonicum* might be caused. To the contrary, the great characteristic of the fused protein of the present invention resides in that a heat-resistant protein derived from a thermophilic bacterium which cannot live in living bodies of mammals and other creatures is used so that even when the fused protein of the present invention is used as an antigen of an immuno-reaction, a nonspecific reaction derived from the fused protein is caused with difficulty.

## BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a detailed view of an expression vector pW6A.

50 Fig. 2 is a detailed view of an expression vector pWF6A.

Fig. 3 is a graph showing the reactivity of a fused protein and a negative specimen.

Fig. 4 is a graph showing the reactivity of a HTLV-I-fused protein and a positive specimen.

Fig. 5 is a graph showing the reactivity of a HTLV-II-fused protein and a positive specimen.

Fig. 6 is a graph showing the reactivity depending on concentration of a HTLV-I-fused protein.

55 Fig. 7 is a graph showing the reactivity depending on concentration of a HTLV-II-fused protein.

Fig. 8 is a graph showing the activity of a fused protein in a supernatant subjected to heat treatment.

Fig. 9 is a graph showing the activity of a fused protein of precipitates subjected to heat treatment.

Fig. 10 is a view showing the activity of a fused protein after heat treatment and purification.

Fig. 11 is a detailed view of an expression vector pW6AK.

Fig. 12 is a view showing the activity of a fused protein after heat treatment and purification.

# DESCRIPTION OF THE PREFERRED EMBODIMENTS

In the following, the present invention is explained in detail.

The DNA sequence coding a heat-resistant protein of the present invention means a DNA sequence coding a protein which is not thermally denatured even at 55 °C or higher, preferably 75 °C or higher. As a specific phenomenon of thermal denaturation, there may be mentioned inactivation or insolubilization of a protein. As the DNA sequence coding a protein which is not thermally denatured at 55 °C or higher, there may be mentioned, for example, a DNA sequence possessed by a thermophilic bacterium which can live at 55 °C or higher. From the properties of an expressed protein and easiness of post-treatment, it is preferred to use a DNA sequence possessed by the so-called highly thermophilic bacterium which can live at 75 °C or higher. As the highly thermophilic bacterium, there may be mentioned, for example, *Thermophilus*, *Sulfolobus*, *Pyrococcus*, *Thermotoga*, *Pyrobaculum*, *Pyrodictum*, *Thermococcus*, *Thermodiscus*, *Metanothermus* and *Metanococcus* (FEMS. MICRO. BIOL. REV., Vol.75, pp.117-124 (1990), ANU. REV. MICROBIOL., Vol.47, pp.627-653 (1993)). As the heat-resistant protein, there may be mentioned, for example, adenyl kinase derived from a *Sulfolobus* bacterium (*Sulfolobus acidocaldalis* Adenylate kinase: Arch. Biochem. Biophys., Vol.207, pp.405-410 (1993)) (hereinafter referred to as "AK" in the specification), DNA polymerase derived from a *Thermophilus* bacterium, ferredoxin derived from a *Pyrococcus* bacterium (*Pyrococcus furiosus* Ferredoxin: Biochemistry, Vol.31, pp.1192-1196 (1992)) (hereinafter referred to as "FDX" in the specification), glucosidase derived from *Pyrococcus furiosus* bacterium (*Pyrococcus furiosus* Glucosidase), rubredoxin derived from *Pyrococcus furiosus* bacterium (*Pyrococcus furiosus* Rubredoxin: Biochemistry, Vol.30, pp.10885-10895 (1991)), glutamate dehydrogenase derived from *Pyrococcus furiosus* bacterium (*Pyrococcus furiosus* Glutamate dehydrogenase: Gene, Vol.132, pp.189-197 (1988)), glyceraldehyde phosphate dehydrogenase derived from *Metanothermus fervidus* bacterium (*Metanothermus fervidus* Glyceraldehyde 3-phosphate dehydrogenase: Gene, Vol.64, p.189-197 (1988)), glutamate synthetase derived from *Metanococcus volatile* bacterium (*Metanococcus volatile* Glutamate synthetase: Res. Microbiol., Vol.140, pp.355-371 (1989)), L-lactate dehydrogenase derived from *Thermotoga maritima* bacterium (*Thermotoga maritima* L-lactate dehydrogenase: Eur. J. Biochem., Vol.216, pp.709-715 (1993)) and elongation factor derived from *Thermococcus celer* bacterium (*Thermococcus celer* Elongation Factor I-alpha: Nucleic acid res. Vol.18, p.3989 (1990)), but the heat-resistant protein coded by the DNA sequence of the present invention is not limited thereby. DNA coding the heat-resistant protein of the present invention can be purified from these highly thermophilic bacteria, but it can be also synthesized based on a known DNA sequence. For synthesis of DNA of the heat-resistant protein, a known technique such as a  $\beta$ -cyanoethylphosphoamidite method (Sinha et al., Nucleic Acids Res., Vol.12, p.4539 (1984)) and a method described in Letsinger, R.L. et al., J. Am. Chem. Soc., vol. 88, p. 5319 (1966) may be suitably used. In Examples each of which is an embodiment of the present invention, DNA's of FDX derived from *Pyrococcus* bacterium and AK derived from *Sulfolobus* bacterium having amino acid sequences shown in SEQ ID NO: 1 and 3, respectively, are synthesized by the  $\beta$ -cyanoethylphosphoamidite method. DNA sequences synthesized are shown in SEQ ID NO: 2 and 4, respectively.

The DNA sequence coding a selected desired protein or peptide of the present invention is not limited to a particular DNA sequence. Any DNA sequence can be used so long as it is a DNA sequence coding a protein or peptide which is desired to be expressed as a fused protein. The present invention is particularly useful when a necessary expression amount of a selected desired protein or peptide can be obtained with difficulty by DNA itself coding said protein or peptide.

The fused DNA sequence of the present invention can be fused by using a known method such as a ligation method and a linker ligation method. When fusion is carried out, the DNA sequence of a selected desired protein or peptide and the DNA sequence of the heat-resistant protein may be fused directly or may be fused indirectly, if necessary. In the case of indirect fusion, a linker sequence is inserted between the DNA sequence coding a desired protein or peptide and the DNA sequence coding the heat-resistant protein. As said linker sequence, there can be used a sequence coding a polypeptide for bonding a desired protein or peptide and the heat-resistant protein to each other and a sequence coding a polypeptide which can be cleaved or digested selectively by a known chemical method or enzymatic method. When the linker sequence is inserted between the DNA sequence coding a desired protein or peptide and the DNA sequence coding the heat-resistant protein, only a selected desired protein or peptide portion can be also purified by, after the fused protein is expressed, cleaving or digesting the linker sequence by using a chemical means such as bromocyan or an enzymatic means such as thrombin or a factor Xa.

In order to express the fused protein of the present invention, a common technique of genetic engineering can be used. For example, the fused DNA sequence of the present invention is inserted into a vector which is suitable for expression, said vector is introduced into a culture host, and expression of the fused protein is induced. After the host is grown by culture or the like, sonication of the host and purification such as a column operation are carried out to obtain a desired fused protein or peptide. Host cells to be used may be any cells such as bacterial cells, eucaryotic cells and mammal cells so long as they are cells which can express a foreign protein or peptide, and there may be mentioned, for example, *Escherichia coli*, yeast, *Bacillus subtilis*, *Baculo* virus and COS cells.

The fused protein of the present invention may be used as such as a fused protein, or a desired protein or peptide portion thereof obtained by separation and purification may be used.

## EXAMPLES

The present invention is described in detail by referring to Reference examples and Examples.

### Example 1 Preparation of FDX-expressing vector pWF6A

By using 8 primers of 53 mer prepared based on a known DNA sequence of *Pyrococcus furiosus* FDX by using a DNA synthesizer (Model 392, trade name, manufactured by PERKIN ELMER Co.), genes of *Pyrococcus furiosus* FDX were synthesized by the assemble PCR (polymerase chain reaction) method. In the assemble PCR method, a Taq polymerase (produced by Toyobo Co.) was used, and the total base number of 248 bp was amplified under conditions of 30 cycles of 94 °C - 1 minute, 55 °C - 1 minute and 72 °C - 1 minute. A NdeI site was added to 5'-end, a restriction enzyme EcoRI was added to 3'-end, and a thrombin-cut site was added to C terminal. This fragment was integrated into the NdeI and EcoRI sites of 4.6 Kb of a pW6A vector prepared from pGEMEX-1 (trade name, produced by Promega Co.) and pGEX-2T (trade name, produced by Pharmacia Biotec Co.) to prepare pWF6A as a vector expressing FDX. A detailed view of pW6A is shown in Fig. 1, and a detailed view of pWF6A is shown in Fig. 2. pWF6A contains, at the NdeI and EcoRI sites, genes of a fused protein comprising 96 amino acids including 67 amino acids derived from FDX, 10 amino acids derived from a thrombin-cleaved site and 19 amino acids derived from multi cloning site of pW6A. The base sequence of the inserted fragment was confirmed by a DNA sequence kit (trade name: Sequenase kit Ver. 2.0, produced by Amersham United States Biochemical Co.). DNA sequence of the FDX inserted into pW6A and amino acids sequence coded by said sequence are shown in SEQ ID NO: 1 and SEQ ID NO: 2, respectively, and DNA sequence of the pW6A is shown in SEQ ID NO: 5. In the sequence table, ATG of the restriction enzyme site NdeI is shown as 1 and sequences up to the stop codon of a multi-cloning site are shown. The expression "\*\*\*\*\*" in the amino acid sequence means the stop codon. pWF6A was introduced into host *Escherichia coli* and then cultured for 2 hours in a medium (hereinafter referred to as "the LB medium" in the specification) containing 1 % of bactotryptone, 0.5 % of yeast extract, 1 % of sodium chloride and 50 µg/ml of ampicillin and having pH 7.5. Thereafter, 1 mM isopropyl thiogalactopyranoside (hereinafter referred to as "IPTG" in the specification) was added thereto, and the mixture was cultured for 2 hours to induce expression. 10 mM Tris-hydrochloride having pH 7.5 and 1 mM ethylenediaminetetraacetic acid (hereinafter abbreviated to as "EDTA" in the specification) (in the following, this buffer is referred to as "a TE buffer" in the specification) were added to the precipitates of *Escherichia coli*, the precipitates were sonicated, and 15 % sodium dodecylsulfate-polyacrylamide gel electrophoresis (hereinafter referred to as "SDS-PAGE") according to the Laemmli method was carried out. By Coomassie brilliant blue staining (hereinafter referred to as "CBB staining" in the specification), a band was confirmed at about 22 Kda, and FDX of *Pyrococcus furiosus* forming a dimer was recognized.

### Example 2 Purification of FDX

pWF6A prepared in Example 1 was introduced into host *Escherichia coli* and then cultured under conditions of using the LB medium at 37 °C. By preculture, a concentration of *Escherichia coli* in a culture broth was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. After the mixture was cultured for 3 hours, centrifugation was carried out to recover *Escherichia coli*. 200 ml of a 50 mM Tris-hydrochloride buffer (hereinafter referred to as "the Tris buffer" in the specification) having pH 8.0 was added to recovered *Escherichia coli*, followed by sonication treatment under ice cooling. After centrifugation, the expressed fused protein was recovered in the supernatant as a soluble component. When this supernatant was subjected heat treatment at 85 °C for 15 minutes, about 80 % of the *Escherichia coli* protein was thermally denatured and precipitated, and 90 % or more of FDX was recovered in the centrifugation supernatant after the heat treatment.

This supernatant was purified by ion exchange using a QFF anion exchange column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with the Tris buffer. When the supernatant was eluted by a column equilibrated buffer containing sodium chloride, FDX was recovered at a concentration of about 0.3 M sodium chloride-eluted fraction. Then, this FDX fraction was purified by using a RESOURCE RPC column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with 20 mM sodium hydroxide. When the fraction was eluted by acetonitrile, purified FDX was recovered at a concentration of about 10 % acetonitrile-eluted fraction.

### Reference example 1 Purification of TRX

pWT8A prepared as a vector expressing TRX in the same manner as in pWF6A prepared in Example 1 was introduced into host *Escherichia coli* and then cultured under conditions of using the LB medium at 37 °C. After the same induction of expression as in Example 1 was carried out, *Escherichia coli* was recovered by centrifugation. An osmotic

shock was given to recovered *Escherichia coli*, and TRX existing at a periplasmic fraction was extracted. Extracted TRX was subjected to first purification by using a RESOURCE RPC column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with 20 mM sodium hydroxide. When TRX was eluted by acetonitrile, TRX was recovered at a concentration of about 10 % to 20 % acetonitrile-eluted fraction. Recovered TRX was dialyzed to 4 M guanidine hydrochloride and then subjected to second purification by using the reverse phase column under the same conditions. Similarly as in the first purification, purified TRX was recovered at a concentration of about 10 % to 20 % acetonitrile-eluted fraction.

#### Example 3 Specificity test of FDX and TRX by the western blotting method

An anti-*Escherichia coli* antibody was supposed as a non-specific reaction substance, and the reactivities of FDX purified in Example 2 and TRX purified in Reference example 1 were examined.

A SDS-solubilized material of *Escherichia coli* DH5 $\alpha$ , a supernatant of *Escherichia coli* DH5 $\alpha$  sonicated and a SDS-solubilized material of *Escherichia coli* to which a pW50 vector (made by Fuji Rebio) was introduced were used as immunogen and immunized to 3 rabbits to prepare the total 9 kinds of the respective anti-*Escherichia coli* rabbit serums. FDX purified in Example 2 and TRX purified in Reference example 1 were subjected to SDS-PAGE according to the Laemmli method and then transferred to nitrocellulose membranes. After blocking the protein portion adsorbed to the nitrocellulose membranes with 1 % skim milk dissolved in PBS, the western blotting method was carried out by using the above 9 kinds of the anti-*Escherichia coli* rabbit serums diluted 500 times, respectively, as primary antibodies, and using a peroxidase (hereinafter referred to as "POD" in the specification)-labeled anti-rabbit antibody as a secondary antibody. For coloring, 4-chloro-1-naphthol and hydrogen peroxide were used. At the portion corresponding to the molecular weight of FDX, no substance reacting with the anti-*Escherichia coli* rabbit antibody was confirmed, but at the portion corresponding to the molecular weight of TRX, among 9 kinds of the anti-*Escherichia coli* rabbit serums, 6 kinds of the serums in which the supernatant of *Escherichia coli* DH5 $\alpha$  sonicated and the SDS-solubilized material of *Escherichia coli* into which the pW50 vector was introduced were used as immunogen were reacted, respectively.

In the same manner as described above, the western blotting method was carried out by 25 samples of human specimen HTLV-III mix panel 204 serums (trade name, produced by Boston Biomedica Co.) diluted 50 times, respectively, as primary antibodies, and using POD-labelled anti-human IgG as a secondary antibody. Reactivities at sites where FDX was transferred was not confirmed, but the reactions of 2 samples among 25 samples at sites where TRX was transferred were confirmed. The results are shown in Table 1.

Table 1

| Specimen No.                | Intensity of reaction (+, -) by western blotting |     |
|-----------------------------|--|-----|
|                             | FDX  | TRX |
| PRP-204-01                  | -  | -   |
| PRP-204-02                  | -  | -   |
| PRP-204-03                  | -  | -   |
| PRP-204-04                  | -  | -   |
| PRP-204-05                  | -  | -   |
| PRP-204-06                  | -  | -   |
| PRP-204-07                  | -  | -   |
| PRP-204-08                  | -  | -   |
| PRP-204-09                  | -  | -   |
| PRP-204-10                  | -  | -   |
| PRP-204-11                  | -  | -   |
| PRP-204-12                  | -  | +   |
| PRP-204-13                  | -  | -   |
| PRP-204-14                  | -  | -   |
| PRP-204-15                  | -  | -   |
| PRP-204-16                  | -  | -   |
| PRP-204-17                  | -  | -   |
| PRP-204-18                  | -  | -   |
| PRP-204-19                  | -  | -   |
| PRP-204-20                  | -  | -   |
| PRP-204-21                  | -  | -   |
| PRP-204-22                  | -  | -   |
| PRP-204-23                  | -  | +   |
| PRP-204-24                  | -  | -   |
| PRP-204-25                  | -  | -   |
| +: positive,<br>-: negative |  |     |

#### Example 4 Specificity test of FDX and TRX by the ELISA method using human specimens

On ELISA plates (produced by Becton Dickinson Co.) were sensitized each 50 µl of 25 µg/ml of FDX purified in Example 2 and TRX purified in Reference example 1, respectively.

After blocking the protein portion adsorbed onto wells of the ELISA plate with 1 % skim milk, a specificity test according to the ELISA method was carried out by using the human specimens produced by Boston Biomedica Co. diluted 500 times used in Example 3 as primary antibodies and POD-labelled anti-human IgG as a secondary antibody. For coloring, ABTS and hydrogen peroxide were used. The measurement results were shown by difference between absorbances at a wavelength of 405 nm and a wavelength of 492 nm (difference between absorbances was described as A405/492 nm). In the reactions with the specimens, whereas there was no specimen exceeding twice of a blank in



the case of FDX, the specimens exceeding twice of a blank were confirmed in 6 samples among 25 samples in the case of TRX. FDX derived from *Pyrococcus furiosus* was different from TRX derived from *Escherichia coli* in that neither nonspecific reaction nor cross reaction derived from *Escherichia coli* was recognized. The results are shown in Fig. 3.

#### 5 Example 5 Expression of FDX-fused HTLV-I p19-fused protein and FDX-fused HTLV-II p19-fused protein

From infected cell lines expressing HTLV-I and HTLV-II, genomic DNA was extracted by the method of Molecular Cloning by J. Sambrook et al. Next, by using a primer to which EcoRI and BamHI sites were added, the PCR method was carried out in the same manner as in Example 1 to obtain about 400 bp of p19DNA fragments in the respective gag regions. These fragments were integrated into pWF6A to prepare pWFIP19 as a vector expressing p19 of HTLV-I and pWFIIIP19 as a vector expressing p19 of HTLV-II. DNA sequences of the FDX-fused HTLV-I p19 and FDX-fused HTLV-II p19 each of which is inserted into the vectors are shown in SEQ ID NO: 6 and 8, respectively, and amino acids sequences coded by said DNA sequences are shown in SEQ ID NO: 7 and 9, respectively. In the same manner as in Example 1, these vectors were introduced into *Escherichia coli*, and expression of the respective fused proteins was induced. Samples for electrophoresis were prepared under the same conditions as in Example 1. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet was transferred to nitrocellulose membranes by the method shown in Example 3. By using an anti-native HTLV-I p19 monoclonal antibody (a GIN-7 antibody, Tanaka, Y. et al., Gann., Vol.74, pp.327 to 330 (1983)) or an anti-native HTLV-II p19 monoclonal antibody as a primary antibody, and a POD-labeled antimouse IgG as a secondary antibody, these were reacted with the fused proteins by the same method as in Example 3 and coloring was carried out by using 4-chloro-1-naphthol and hydrogen peroxide, expression of the fused proteins reacting with the respective monoclonal antibodies corresponding to the respective fused proteins was recognized. These fused proteins gave a band at about 34 Kda which was the same position as that of the CBB-stained gels. The expression amounts of the FDX-fused HTLV-I p19 antigen and the FDX-fused HTLV-II p19 antigen were increased by several hundreds times as compared with the case where the p19 antigen of HTLV-I and the p19 antigen of HTLV-II were expressed directly.

#### Example 6 Expression of FDX-fused HTLV-I p20E(gp21)-fused protein and HTLV-II p20E(gp21)-fused protein

By the same method as in Example 5, by using DNA of cells infected with HTLV-I and HTLV-II, about 500 bp of p20E(gp21) DNA fragments in the respective env regions were obtained by the PCR method. These DNA fragments were integrated into EcoRI and BamHI of pWF6A prepared in Example 1 to prepare pWFIE1 as a vector expressing p20E of HTLV-I and pWFIIIE10 as a vector expressing p20E of HTLV-II. DNA sequences of the FDX-fused HTLV-I p20E and FDX-fused HTLV-II p20E each of which is inserted into the vectors are shown in SEQ ID NO: 10 and 12, respectively, and amino acids sequences coded by said DNA sequences are shown in SEQ ID NO: 11 and 13, respectively. These vectors were introduced into *Escherichia coli*, and expression of a FDX-fused HTLV-I p20E-fused protein (hereinafter referred to as "FDX-20(I)" in the specification) and a FDX-fused HTLV-II p20E-fused protein (hereinafter referred to as "FDX-20(II)" in the specification) was induced under the same conditions as in Example 1. In the same manner as in Example 1, *Escherichia coli* was sonicated. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet of gel was transferred to nitrocellulose membranes at 120 mA for 3 hours. After blocking the protein portion adsorbed to the nitrocellulose membranes with a phosphate buffer containing 1 % of BSA (bovine serum albumin), 1 µg/ml of an anti-p20E(gp21) monoclonal antibody (F-10, Sugamura, K. et al., J. Immunol., Vol.132, pp.3180 to 3184 (1984)) reacting with p20E(gp21) antigens of native HTLV-I and HTLV-II was reacted with the fused proteins at room temperature for 1 hour, and then reacted with a POD-labeled anti-mouse IgG at room temperature for 1 hour. Subsequently, when coloring was carried out by using 4-chloro-1-naphthol and hydrogen peroxide, expression of fused proteins reacting with the anti-p20E(gp21) monoclonal antibody corresponding to the respective fused proteins was recognized. These fused proteins gave a band at about 32 Kda which was the same position as that of the CBB-stained gels.

The expression amounts of FDX-20(I) and FDX-20(II) were increased by several hundreds times as compared with the case where p20E of HTLV-I and p20E of HTLV-II were expressed directly.

#### 50 Example 7 Purification of FDX-20(I)- and FDX-20(II)-fused proteins

pWFIE1 and pWFIIIE10 prepared in Example 6 were introduced into host *Escherichia coli*, respectively, and then cultured under conditions of using the LB medium at 37 °C. By preculture, a concentration of *Escherichia coli* in culture broths was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. Three hours after IPTG was added, centrifugation was carried out to recover *Escherichia coli*. 200 ml of a 50 mM Tris-hydrochloride buffer containing 1 % Triton X 100 (trade name, produced by Rohm & Haas Co.) and 2 M urea with pH 8.0 was added to recovered *Escherichia coli*, followed by sonication treatment under ice cooling. Centrifugation was carried out to recover insoluble materials (inclusion bodies). The inclusion bodies

were solubilized by using a 4 M guanidine hydrochloride-10 mM dithiothreitol (hereinafter referred to as "DTT" in the specification) solution. The solubilized bodies were purified by a RESOURCE RPC column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with 20 % acetonitrile and 20 mM sodium hydroxide. When the bodies were eluted by acetonitrile, purified FDX-20(I)- and FDX-20(II)-fused proteins were recovered at a concentration of about 30 to 40 % acetonitrile-eluted fractions, respectively.

#### Reference example 2 Purification of TRX-fused HTLV-I p20E-fused protein and TRX-fused HTLV-II p20E-fused protein

In the same manner as in Example 6, p20E(gp21) in an env region of HTLV-I or HTLV-II was introduced into the TRX-expressing vector pWT8A prepared in Reference example 1 to prepare pWTIE1 and pWTIE10, followed by expression. In the same manner as in Example 7, by the purification method using a RESOURCE RPC column (trade name, manufactured by Pharmacia Biotec Co.), a TRX-fused HTLV-I p20E-fused protein (hereinafter referred to as "TRX-20(I)" in the specification) and a TRX-fused HTLV-II p20E-fused protein (hereinafter referred to as "TRX-20(II)" in the specification) were purified.

#### Example 8 Reactivity test of fused proteins

##### (1) Test by the western blotting method

By using FDX-20(I) and FDX-20(II) purified in Example 7 and TRX-20(I) and TRX-20(II) purified in Reference example 2, reactivities with human HTLV specimens in the western blotting method were compared.

In the same manner as in Example 3, the western blotting method was carried out by using the human specimen HTLV-I/II mix panel produced by Boston Biomedica Co. diluted 50 times as primary antibodies and POD-labelled human IgG as a secondary antibody. FDX-20(I) and FDX-20(II), and TRX-20(I) and TRX-20(II) were reacted with the same specimens, respectively. The results are shown in Table 2.

Table 2

| Specimen No.  | Intensity of reaction (+, -) by western blotting |           |            |            |
|---|--|-----------|------------|------------|
|   | FDX-20(I)  | TRX-20(I) | FDX-20(II) | TRX-20(II) |
| PRP-204-01  | +  | +         | +          | +          |
| PRP-204-02  | -  | -         | -          | -          |
| PRP-204-03  | +  | +         | +          | +          |
| PRP-204-04  | -  | -         | +          | +          |
| PRP-204-05  | +  | +         | -          | -          |
| PRP-204-06  | -  | -         | -          | +          |
| PRP-204-07  | +  | +         | +          | +          |
| PRP-204-08  | -  | -         | -          | -          |
| PRP-204-09  | +  | +         | -          | -          |
| PRP-204-10  | +  | +         | +          | +          |
| PRP-204-11  | +  | +         | +          | +          |
| PRP-204-12  | ++   | ++        | ++         | ++         |
| PRP-204-13  | +  | +         | +          | +          |
| PRP-204-14  | -  | -         | +          | +          |
| PRP-204-15  | +  | +         | +          | +          |
| PRP-204-16  | -  | -         | +          | +          |
| PRP-204-17  | +  | +         | +          | +          |
| PRP-204-18  | +  | +         | +          | +          |
| PRP-204-19  | +  | +         | -          | -          |
| PRP-204-20  | -  | -         | -          | -          |
| PRP-204-21  | +  | +         | +          | +          |
| PRP-204-22  | +  | +         | +          | +          |
| PRP-204-23  | +  | +         | +          | +          |
| PRP-204-24  | +  | +         | +          | +          |
| PRP-204-25  | +  | +         | +          | +          |
| +: positive,<br>++: strongly positive,<br>-: negative |  |           |            |            |

## (2) Comparison by the ELISA method

On ELISA plates (produced by Becton Dickinson Co.) were sensitized each 50  $\mu$ l of FDX-20(I) and FDX-20(II) purified in Example 7 and TRX-20(I) and TRX-20(II) purified in Reference example 2 at a concentration of 3  $\mu$ g/ml, respectively.

The ELISA method was carried out by using these ELISA plates and using the human specimens produced by Boston Biomedica Co. diluted 500 times as primary antibodies and POD-labelled anti-human IgG as a secondary antibody in the same manner as in Example 4. FDX-20(I) and FDX-20(II), and TRX-20(I) and TRX-20(II) were reacted with the same specimens. The results are shown in Fig. 4 and Fig. 5.

## (3) Test of dependency on concentration by the ELISA method

In order to examine reactivities to the anti-p20E(gp21) monoclonal antibody and a negative serum, 10 µg/ml to 1/2 dilution series of FDX-20(I) and FDX-20(II) purified in Example 7 and TRX-20(I) and TRX-20(II) purified in Reference example 2 were prepared, respectively, and ELISA plates (produced by Becton Dickinson Co.) were sensitized with each 50 µl thereof.

The ELISA method was carried out by using these ELISA plates and using the anti-p20E(gp21) monoclonal antibody diluted 500 times as a primary antibody and POD-labelled anti-mouse IgG as a secondary antibody. With respect to a negative serum, the ELISA method was carried out in the same manner as in Example 4. There was no difference in reactivity to the monoclonal antibody, and the FDX-fused proteins in both cases of HTLV-I and HTLV-II had lower reactivities to the negative serum. The results are shown in Fig. 6 and Fig. 7.

Reference example 3 Preparation of protein in which GST and *Treponema pallidum* 15Kda antigen are fused

From syphilis bacteria (Nichols strain from *Treponema pallidum*) purified from syphilis bacteria-subcultured rabbit testicles, genomic DNA was extracted. By using the extracted DNA as a template, a primer was produced based on the known DNA sequences by using a DNA synthesizer (Model 392, trade name, produced by PERKIN ELMER Co.). By using the primer, about 370 bp of a DNA fragment coding a surface antigen of 15 Kda (hereinafter referred to as "Tp15" in the specification) of *Treponema pallidum* (hereinafter referred to as "Tp" in the specification) was amplified with a thermal cycler (Model PJ1000, trade name, produced by PERKIN ELMER Co.). This DNA fragment was integrated into an EcoRI site of a GST-expressing type vector pWG6A in which DNA sequence of GST had been inserted into pW6A to obtain a vector pWGTP15 expressing a protein in which GST and Tp15 were fused (hereinafter referred to as "GST-15" in the specification). DNA sequence of the GST-15 inserted into the vector is shown in SEQ ID NO: 14 and amino acids sequence coded by said DNA sequence is shown in SEQ ID NO: 15. In the same manner as in Example 1, the vector was introduced into *Escherichia coli*, and expression of GST-15 was induced. A sample for electrophoresis was prepared under the same conditions as in Example 1. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet was transferred to a nitrocellulose membrane by the method shown in Example 3. By using an anti-Tp15 monoclonal antibody as a primary antibody and a POD-labeled mouse IgG as a secondary antibody, these were reacted in the same method as in Example 3 and coloring was carried out by using 4-chloro-1-naphthol and hydrogen peroxide, a band was given at about 42 Kda which was the same position as that of the CBB-stained gel.

## Reference example 4 Preparation of protein in which TRX and Tp15 are fused

A DNA fragment of Tp15 amplified in Reference example 3 was integrated into an EcoRI site of the TRX-expressing type vector pWT8A in which DNA sequence of TRX had been inserted into pW6A to obtain a vector pWTP15 expressing a protein in which TRX and Tp15 were fused (hereinafter referred to as "TRX-15" in the specification). DNA sequence of the TRX-15 inserted into the vector is shown in SEQ ID NO: 16 and amino acids sequence coded by said DNA sequence is shown in SEQ ID NO: 17. In the same manner as in Example 1, the vector was introduced into *Escherichia coli*, and expression of TRX-15 was induced. A sample for electrophoresis was prepared under the same conditions as in Example 1. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet was transferred to a nitrocellulose membrane by the method shown in Example 3. By using an anti-Tp15 monoclonal antibody as a primary antibody and a POD-labeled mouse IgG as a secondary antibody, these were reacted in the same method as in Example 3 and coloring was carried out by using 4-chloro-1-naphthol and hydrogen peroxide, a band was given at about 27 Kda which was the same position as that of the CBB-stained gel.

## Example 9 Preparation of protein in which FDX and Tp15 are fused

A DNA fragment of Tp15 amplified in Reference example 3 was integrated into an EcoRI, BamHI site of the FDX-expressing type vector pWF6A prepared in Example 1 to obtain a vector pWFTp15 expressing a protein in which FDX and Tp15 were fused (hereinafter referred to as "FDX-15" in the specification). DNA sequence of the FDX-15 inserted into the vector is shown in SEQ ID NO: 18 and amino acids sequence coded by said DNA sequence is shown in SEQ ID NO: 19. In the same manner as in Example 1, the vector was introduced into *Escherichia coli*, and expression of FDX-15 was induced. A sample for electrophoresis was prepared under the same conditions as in Example 1. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet was transferred to a nitrocellulose membrane by the method shown in Example 3. By using an anti-Tp15 monoclonal antibody as a primary antibody and a POD-labeled mouse IgG as a secondary antibody, these were reacted in the same method as in Example 3 and coloring was carried out by using 4-chloro-1-naphthol and hydrogen

peroxide, a band was given at about 30 Kda which was the same position as that of the CBB-stained gel.

#### Example 10 Heat resistance test of FDX-15, GST-15 and TRX-15

5 The vectors expressing FDX-15, GST-15 and TRX-15 prepared in Example 9, Reference example 3 and Reference example 4 were introduced into host *Escherichia coli* and then cultured under conditions of using 1 liter of the LB medium at 37 °C, respectively. By preculture, a concentration of *Escherichia coli* in culture broths was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. After the cells were recovered by centrifugation, 200 ml of the Tris buffer was added to the cells. After sonication treatment under ice cooling, fused proteins were recovered in the centrifugation supernatants, respectively. 800  $\mu$ l of these proteins were taken, respectively, and shaken for 13 minutes in water bath at 40 °C, 50 °C, 60 °C, 70 °C and 80 °C. The respective samples were centrifuged and then separated into supernatants and precipitates, and analysis was carried out by SDS-PAGE and the western blotting method. As a blocking agent of the western blotting method, 1 % skim milk dissolved in PBS was used, and as a primary antibody, an anti-TP rabbit antibody was used. As a secondary antibody, a POD-labelled anti-rabbit antibody was used, and as a coloring agent, 4-chloro-1-naphthol and hydrogen peroxide were used. The result of coloring of western blotting was confirmed by a densitometer. The results are shown in Fig. 8 and Fig. 9. Precipitates of TRX-15 and GST-15 were partially generated at 40 °C by thermal denaturation, about 80 % of TRX-15 and GST-15 were precipitated at 60 °C, and about 100 % of them were precipitated at 70 °C. Almost no precipitate by thermal denaturation of FDX-15 was generated at 40 °C to 80 °C, and even at 80 °C, about 100 % of FDX-15 existed in the supernatant.

#### Example 11 Purification of FDX-15 by heat treatment

pWFTp15 prepared in Example 9 was introduced into host *Escherichia coli* and then cultured under conditions of using 1 liter of the LB medium at 37 °C. By preculture, a concentration of *Escherichia coli* in culture broths was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. The cells were recovered by centrifugation. 200 ml of the Tris buffer was added to the cells, and the cells were sonicated to recover FDX-15 in the centrifugation supernatant. Then, by using a hot plate and a water bath, heat treatment at 70 °C for 10 minutes was carried out to recover FDX-15 in the centrifugation supernatant. The supernatant subjected to heat treatment was purified by a QFF anion exchange column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with the Tris buffer. When the supernatant was eluted by a column equilibrated buffer containing sodium chloride, FDX-15 was recovered at a concentration of about 0.3 M to 0.4 M sodium chloride-eluted fraction. Then, 10 mM DTT was added to the QFF recovered fraction, and the mixture was purified by using a RESOURCE RPC column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with a 20 mM sodium hydroxide solution. When the mixture was eluted by acetonitrile, FDX-15 was recovered at a concentration of about 20 % to 25 % acetonitrile-eluted fraction. This reverse phase recovered fraction was concentrated by Centriprep (trade name, manufactured by Amicon Inc.), and the concentrate was subjected to gel filtration by a Superdex 200 column (trade name, manufactured by Pharmacia Biotec Co.). When the filtrate was eluted by a buffer containing 6 M urea, 0.5 M sodium chloride and 20 mM Tris-hydrochloride having pH 8.0, purified FDX-15 was recovered at a molecular weight of about 50,000. By heat treatment at 60 °C, about 80 % of the *Escherichia coli* protein was precipitated by thermal denaturation, but even at 70 °C, almost 100 % of FDX-15 was recovered in the supernatant, and the purification degree was raised by about 5 times only by heat treatment.

Further, GST-15 obtained by introducing pWGTp15 prepared in Reference example 3 into host *Escherichia coli*, carrying out induction and expression operations in the same manner therein and carrying out purification by a common column operation without carrying out heat treatment and FDX-15 purified by heat treatment were subjected to the western blotting method in the same manner as in Example 10 by using an anti-Tp rabbit antibody. It was shown that even though purification by heat treatment was carried out, FDX-15 retained reactivity. The results are shown in Fig. 10.

#### Example 12 Preparation of AK-expressing vector pW6AK

50 By using 16 primers of 53 mer prepared based on a known DNA sequence of AK derived from a *Sulfolobus* bacterium by using a DNA synthesizer (manufactured by Perkin Elmer Co.), genes of *Sulfolobus acidocaldarius* AK were synthesized by the assemble PCR method. In the assemble PCR method, a Taq polymerase (produced by Toyobo Co.) was used, and the total base number of 630 bp was amplified under conditions of 30 cycles of 94 °C - 1 minute, 55 °C - 1 minute and 72 °C - 1 minute. A NdeI site was added to 5'-end, a restriction enzyme EcoRI was added to 3'-end, and a thrombin-cut site was added to C terminal. This fragment was integrated into the NdeI and EcoRI sites of 4.6 Kb of a pW6A vector prepared from pGEMEX-1 (trade name, produced by Promega Co.) and pGEX-2T (trade name, produced by Pharmacia Biotec Co.) to prepare pW6AK as a vector expressing AK. A detailed view of pW6AK is shown in Fig. 11. pW6AK contains genes of a fused protein comprising 223 amino acids including 194 amino acids derived from AK, 10

amino acids derived from a thrombin-cleaved site and 19 amino acids derived from multi cloning site of pW6A, at the NdeI and EcoRI sites. The base sequence of the inserted fragment was confirmed by a DNA sequence kit (trade name: Sequenase kit Ver. 2.0, produced by Amersham United States Biochemical Co.). DNA sequence of the AK inserted into the pW6A is shown in SEQ ID NO: 3 and amino acids sequence coded by said DNA sequence is shown in SEQ ID NO: 4. pW6AK was introduced into host *Escherichia coli* and then cultured for 2 hours in the LB medium. Thereafter, 1 mM IPTG was added thereto, and the mixture was cultured for 2 hours to induce expression. The TE buffer were added to the precipitates of *Escherichia coli*, the precipitates were sonicated, and 15 % SDS-PAGE according to the Laemmli method was carried out. By CBB staining, a band was confirmed at about 40 Kda.

#### Example 13 Purification of AK

pW6AK prepared in Example 12 was introduced into host *Escherichia coli* and then cultured under conditions of using the LB medium at 37 °C. By preculture, a concentration of *Escherichia coli* in culture broth was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. After the mixture was cultured for 3 hours, centrifugation was carried out to recover *Escherichia coli*. 200 ml of the Tris buffer was added to recover *Escherichia coli*, followed by sonication treatment under ice cooling. After centrifugation, the expressed fused protein was recovered in the supernatant as a soluble component. When this supernatant was subjected to heat treatment at 65 °C for 10 minutes, about 70 % of the *Escherichia coli* protein was thermally denatured and precipitated, and 80 % or more of AK was recovered in the centrifugation supernatant after the heat treatment.

This supernatant was purified by a Hydroxy apatite column (manufactured by Bio-rad Lab.) equilibrated with the Tris buffer. When the supernatant was eluted by a sodium phosphate buffer, AK was recovered at a concentration of about 0.2 M sodium phosphate-eluted fraction. Then, this AK fraction was purified by gel filtration using a Superdex 200 26/60 column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with a buffer containing 6 M urea, 0.5 M sodium chloride and 20 mM Tris-hydrochloride having pH 9.4. At a fraction of a molecular weight being about 20,000, purified AK was recovered.

#### Example 14 Preparation of protein in which AK and Tp15 are fused

A DNA fragment of Tp15 amplified in Reference example 3 was integrated into the AK-expressing type vector pW6AK prepared in Example 12 to obtain a vector pW6AKTp15 expressing a protein in which AK and Tp15 were fused (hereinafter referred to as "AK-15" in the specification). DNA sequence of the AK-15 inserted into the vector is shown in SEQ ID NO: 20 and amino acids sequence coded by said DNA sequence is shown in SEQ ID NO: 21. In the same manner as in Example 1, the vector was introduced into *Escherichia coli*, and expression of AK-15 was induced. A sample for electrophoresis was prepared under the same conditions as in Example 1. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet was transferred to a nitrocellulose membrane by the method shown in Example 3. By using an anti-Tp15 monoclonal antibody as a primary antibody and a POD-labeled mouse IgG as a secondary antibody, these were reacted in the same method as in Example 3 and coloring was carried out by using 4-chloro-1-naphthol and hydrogen peroxide, a band was given at about 40 Kda which was the same position as that of the CBB-stained gel.

#### Example 15 Purification of AK-15 by heat treatment

pWAKTp15 prepared in Example 14 was introduced into host *Escherichia coli* and then cultured under conditions of using 1 liter of the LB medium at 37 °C. By preculture, a concentration of *Escherichia coli* in culture broth was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. The cells were recovered by centrifugation. 200 ml of a 50 mM glycine-sodium hydroxide buffer having pH 10.0 was added to the cells, and the cells were sonicated to recover AK-15 in the centrifugation supernatant. Then, by using a hot plate, heat treatment at 60 °C for 10 minutes was carried out to recover AK-15 in the centrifugation supernatant. The supernatant subjected to heat treatment was dialyzed to a 4 M urea-50 mM sodium acetate buffer having pH 6.0 and then purified by a SFF cation exchange column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with said buffer. When the supernatant was eluted by a column equilibrated buffer containing sodium chloride, AK-15 was recovered at a concentration of about 0.2 M to 0.4 M sodium chloride-eluted fraction. The recovered AK-15 fraction was purified by gel filtration using a Superdex 200 26/60 column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with a buffer containing 6 M urea, 0.5 M sodium chloride and 20 mM Tris-hydrochloride having pH 9.4. At a fraction of a molecular weight being about 40,000, purified AK-15 was recovered.

When the western blotting method was carried out in the same manner as in Example 1 by using an anti-Tp rabbit antibody, it was shown that even though purification by heat treatment was carried out, AK-15 retained reactivity. The results are shown in Fig. 12.

# EP 0 781 848 A2

According to the present invention, a fused DNA sequence having more excellent operatability and productivity than those of a conventional DNA sequence coding a fused protein, a fused protein expressed from said fused DNA sequence, and a method for expressing the fused protein by using said DNA sequence.

5 1 RAW SEQUENCE LISTING  
2 PATENT APPLICATION  
3  
4 SEQUENCE LISTING  
5  
10 6 (1) GENERAL INFORMATION:  
7 (i) APPLICANT: Nobuyuki FUJII et al.  
8 (ii) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
9 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
10 (iii) NUMBER OF SEQUENCES:  
11 (iv) CORRESPONDENCE ADDRESS:  
15 12 (A) ADDRESSEE: c/o FUJIREBIO INC., 7-1  
13 (B) STREET: Nishi-shinjuku 2-chome  
14 (C) CITY: Shinjuku-ku  
15 (D) STATE: Tokyo  
16 (E) COUNTRY: JAPAN  
17 (F) ZIP: 163-07  
20 18 (v) COMPUTER READABLE FORM:  
19 (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
20 (B) COMPUTER: IBM Compatible  
21 (C) OPERATING SYSTEM: MS-DOS v.5  
22 (D) SOFTWARE: Word Perfect 5.1  
23 (vi) CURRENT APPLICATION DATA:  
25 24 (A) APPLICATION NUMBER:  
25 (B) FILING DATE:  
26 (vii) PRIOR APPLICATION DATA:  
27 (A) APPLICATION NUMBER: JP 352225/1995  
28 (B) FILING DATE: 28-DEC-1995  
30 29 (viii) ATTORNEY/AGENT INFORMATION:  
30 (A) NAME:  
31 (B) REGISTRATION NUMBER:  
32 (C) REFERENCE/DOCKET NUMBER:  
33 (ix) TELECOMMUNICATION INFORMATION:  
34 (A) TELEPHONE:  
35 35 (B) TELEFAX:  
36  
37 (2) INFORMATION FOR SEQ ID NO: 1:  
38 (i) SEQUENCE CHARACTERISTICS:  
39 (A) LENGTH: 291 nucleic acids  
40 40 (B) TYPE: nucleic acid  
41 (C) STRANDEDNESS: double strand  
42 (D) TOPOLOGY: linear  
43 (ii) MOLECULE TYPE: other nucleic acid  
44 (vi) ORIGINAL SOURCE:  
45 45 (A) ORGANISM: synthesized  
46 (x) PUBLICATION INFORMATION:  
45 47 (A) AUTHORS: Nobuyuki FUJII et al.  
48 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
49 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
50 (K) RELEVANT RESIDUES IN SEQ ID NO:1: FROM 1 to 291  
51  
52 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:  
53  
54 ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC  
55 30 60  
56 CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCCAACCAA GGTAGAGGTT  
57 90 120  
58 ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT

59 150 180  
60 GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGGATC GGAATTCGTC  
61 210 240  
5 62 GACCTCGAGG GATCCGGGCC CTCTAGATGC GGCCGCATGC ATGGTACCTA A  
63 270  
64  
65 (2) INFORMATION FOR SEQ ID NO: 2:  
66 (i) SEQUENCE CHARACTERISTICS:  
10 67 (A) LENGTH: 97 amino acids  
68 (B) TYPE: amino acid  
69 (D) TOPOLOGY: linear  
70 (ii) MOLECULE TYPE:  
71 (A) DESCRIPTION: protein  
72 (vi) ORIGINAL SOURCE:  
15 73 (A) ORGANISM: recombinant  
74 (x) PUBLICATION INFORMATION:  
75 (A) AUTHORS: Nobuyuki FUJII et al.  
76 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
77 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
78 (K) RELEVANT RESIDUES IN SEQ ID NO:2: FROM 1 to 97  
20 79  
80 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:  
81  
82 Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser  
83 1 20  
25 84 Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val  
85 21 40  
86 Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser  
87 41 60  
88 Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Val  
89 61 80  
30 90 Asp Leu Glu Gly Ser Gly Pro Ser Arg Cys Gly Arg Met His Gly Thr \*\*\*  
91 81  
92  
93 (2) INFORMATION FOR SEQ ID NO: 3:  
94 (i) SEQUENCE CHARACTERISTICS:  
35 95 (A) LENGTH: 672 nucleic acids  
96 (B) TYPE: nucleic acid  
97 (C) STRANDEDNESS: double strand  
98 (D) TOPOLOGY: linear  
99 (ii) MOLECULE TYPE: other nucleic acid  
100 (vi) ORIGINAL SOURCE:  
101 (A) ORGANISM: synthesized  
40 102 (x) PUBLICATION INFORMATION:  
103 (A) AUTHORS: Nobuyuki FUJII et al.  
104 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
105 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
106 (K) RELEVANT RESIDUES IN SEQ ID NO:3: FROM 1 to 672  
107  
45 108 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:  
109  
110 ATGAAGATTG GTATTGTAAC TGGTATCCCT GGTGTAGGGA AAAGTACTGT CTTGGCTAAA  
111 30 60  
112 GTTAAAGAGA TATTGGATAA TCAAGGTATA AATAACAAGA TCATAAATTA TGGAGATTTT  
113 90 120  
50 114 ATGTTAGCAA CAGCATTAAG ATTAGGCTAT GCTAAAGATA GAGACGAAAT GAGAAAATTA  
115 150 180  
116 TCTGTAGAAA AGCAGAAGAA ATTGCAGATT GATGCGGCTA AAGGTATAGC TGAAGAGGCA

55



```

117.                                     210                                     240
118 AGAGCAGGTG GAGAAGGATA TCTGTTTCATA GATACGCACG CTGTGATACG TACACCCTCT
119                                     270                                     300
5  120 GGATATTTAC CTGGTTTACC GTCAGATATA ATTACAGAAA TAAATCCGTC TGTTATCTTT
121                                     330                                     360
122 TTAAGTGAAG CTGATCCTAA GATAATATTA TCAAGGCAAA AGAGAGATAC AACAAGGAAT
123                                     390                                     420
124 AGAAATGATT ATAGTGACGA ATCAGTTATA TTAGAAACCA TAAACTTCGC TAGATATGCA
125                                     450                                     480
10 126 GCTACTGCTT CTGCAGTATT AGCCGGTTCT ACTGTTAAGG TAATTGTAAA CGTGGAAGGA
127                                     510                                     540
128 GATCCTAGTA TAGCAGCTAA TGAGATAATA AGGTCTATGA AGGGTGGTTC TTCTCTGGTT
129                                     570                                     600
130 CCGCGTGGAC TGGAATTCGT CGACCTCGAG GGATCCGGGC CCTCTAGATG CGGCCGCATG
131                                     630                                     660
15 132 CATGGTACCT AA
133
134 (2) INFORMATION FOR SEQ ID NO: 4:
135 (i) SEQUENCE CHARACTERISTICS:
136 (A) LENGTH: 224 amino acids
20 137 (B) TYPE: amino acid
138 (D) TOPOLOGY: linear
139 (ii) MOLECULE TYPE:
140 (A) DESCRIPTION: protein
141 (vi) ORIGINAL SOURCE:
142 (A) ORGANISM: recombinant
25 143 (x) PUBLICATION INFORMATION:
144 (A) AUTHORS: Nobuyuki FUJII et al.
145 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID
146 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
147 (K) RELEVANT RESIDUES IN SEQ ID NO:4: FROM 1 to 224
148
30 149 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
150
151 Met Lys Ile Gly Ile Val Thr Gly Ile Pro Gly Val Gly Lys Ser Thr Val Leu Ala Lys
152 1 20
153 Val Lys Glu Ile Leu Asp Asn Gln Gly Ile Asn Asn Lys Ile Ile Asn Tyr Gly Asp Phe
154 21 40
35 155 Met Leu Ala Thr Ala Leu Lys Leu Gly Tyr Ala Lys Asp Arg Asp Glu Met Arg Lys Leu
156 41 60
157 Ser Val Glu Lys Gln Lys Lys Leu Gln Ile Asp Ala Ala Lys Gly Ile Ala Glu Glu Ala
158 61 80
159 Arg Ala Gly Gly Glu Gly Tyr Leu Phe Ile Asp Thr His Ala Val Ile Arg Thr Pro Ser
40 160 81 100
161 Gly Tyr Leu Pro Gly Leu Pro Ser Asp Ile Ile Thr Glu Ile Asn Pro Ser Val Ile Phe
162 101 120
163 Leu Leu Glu Ala Asp Pro Lys Ile Ile Leu Ser Arg Gln Lys Arg Asp Thr Thr Arg Asn
164 121 140
165 Arg Asn Asp Tyr Ser Asp Glu Ser Val Ile Leu Glu Thr Ile Asn Phe Ala Arg Tyr Ala
45 166 141 160
167 Ala Thr Ala Ser Ala Val Leu Ala Gly Ser Thr Val Lys Val Ile Val Asn Val Glu Gly
168 161 180
169 Asp Pro Ser Ile Ala Ala Asn Glu Ile Ile Arg Ser Met Lys Gly Gly Ser Ser Leu Val
170 181 200
171 Pro Arg Gly Leu Glu Phe Val Asp Leu Glu Gly Ser Gly Pro Ser Arg Cys Gly Arg Met
50 172 201 220
173 His Gly Thr ***
174 221

```

55

175  
176 (2) INFORMATION FOR SEQ ID NO: 5:  
177 (i) SEQUENCE CHARACTERISTICS:  
5 178 (A) LENGTH: 4557 nucleic acids  
179 (B) TYPE: nucleic acid  
180 (C) STRANDEDNESS: double strand  
181 (D) TOPOLOGY: circular  
182 (ii) MOLECULE TYPE: other nucleic acid  
183 (vi) ORIGINAL SOURCE:  
10 184 (A) ORGANISM: *E. coli*  
185 (B) STRAIN: BL21(DE3)  
186 (x) PUBLICATION INFORMATION:  
187 (A) AUTHORS: Nobuyuki FUJII et al.  
188 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
15 189 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
190 (K) RELEVANT RESIDUES IN SEQ ID NO:5: FROM 1 to 4557  
191  
192 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:  
193  
194 ATGGCTAGCG AATTCGTCGA CCTCGAGGGA TCCGGGCCCT CTAGATGCGG CCGCATGCAT  
20 195 30 60  
196 GGTACCTAAC TAACTAAGCT TGAGTATTCT ATAGTGTCAC CTAAATCCCA GCTTGATCCG  
197 90 120  
198 GCTGCTAACA AAGCCCGAAA GGAAGCTGAG TTGGCTGCTG CCACCGCTGA GCAATAACTA  
199 150 180  
200 GCATAACCCC TTGGGGCCTC TAAACGGGTC TTGAGGGGTT TTTTGCTGAA AGGAGGAACT  
25 201 210 240  
202 ATATCCGGAT AACCTGGCGT AATAGCGAAG AGGCCCGCAC CGAATTAATT CATCGTGAAT  
203 270 300  
204 GACTGACGAT CTGCCCTCGCG CGTTTCGGTG ATGACGGTGA AAACCTCTGA CACATGCAGC  
205 330 360  
206 TCCCGGAGAC GGTACAGCT TGTCTGTAAG CGGATGCCGG GAGCAGACAA GCCCGTCAGG  
30 207 390 420  
208 GCGCGTCAGC GGGTGTGGC GGGTGTGGG GCGCAGCCAT GACCCAGTCA CGTAGCGATA  
209 450 480  
210 GCGGAGTGTA TAATTCTTGA AGACGAAAGG GCCTCGTGAT ACGCCTATTT TTATAGGTTA  
211 510 540  
212 ATGTCATGAT AATAATGGTT TCTTAGACGT CAGGTGGCAC TTTTCGGGGA AATGTGCGCG  
35 213 570 600  
214 GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGAGACAAT  
215 630 660  
216 AACCCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTCC  
217 690 720  
218 GTGTCGCCCT TATTCCTTTT TTTGCGGCAT TTTGCCTTCC TGTTTTTGCT CACCCAGAAA  
40 219 750 780  
220 CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT TACATCGAAC  
221 810 840  
222 TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT TTTCCAATGA  
223 870 900  
224 TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTGTGAC GCCGGGCAAG  
45 225 930 960  
226 AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTGAGTAC TCACCACTCA  
227 990 1020  
228 CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA  
229 1050 1080  
50 230 TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA  
231 1110 1140  
232 CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC

55

EP 0 781 848 A2

|    |     |            |            |            |            |            |             |
|----|-----|------------|------------|------------|------------|------------|-------------|
|    | 233 |            |            | 1170       |            |            | 1200        |
|    | 234 | TGAATGAAGC | CATACCAAAC | GACGAGCGTG | ACACCACGAT | GCCTGCAGCA | ATGGCAACAA  |
|    | 235 |            |            | 1230       |            |            | 1260        |
| 5  | 236 | CGTTGCGCAA | ACTATTAAC  | GGCGAACTAC | TTACTCTAGC | TTCCCGGCAA | CAATTAATAG  |
|    | 237 |            |            | 1290       |            |            | 1320        |
|    | 238 | ACTGGATGGA | GGCGGATAAA | GTTGCAGGAC | CACTTCTGCG | CTCGGCCCTT | CCGGCTGGCT  |
|    | 239 |            |            | 1350       |            |            | 1380        |
|    | 240 | GGTTTATTCG | TGATAAATCT | GGAGCCGGTG | AGCGTGGGTC | TCGCGGTATC | ATTGCAGCAC  |
|    | 241 |            |            | 1410       |            |            | 1440        |
| 10 | 242 | TGGGGCCAGA | TGGTAAGCCC | TCCCGTATCG | TAGTTATCTA | CACGACGGGG | AGTCAGGCAA  |
|    | 243 |            |            | 1470       |            |            | 1500        |
|    | 244 | CTATGGATGA | ACGAAATAGA | CAGATCGCTG | AGATAGGTGC | CTCACTGATT | AAGCATTGGT  |
|    | 245 |            |            | 1530       |            |            | 1560        |
|    | 246 | AACTGTCAGA | CCAAGTTTAC | TCATATATAC | TTTAGATTGA | TTTAAACTTT | CATTTTTTAAT |
|    | 247 |            |            | 1590       |            |            | 1620        |
| 15 | 248 | TTAAAAGGAT | CTAGGTGAAG | ATCCTTTTTG | ATAATCTCAT | GACCAAAATC | CCTTAACGTG  |
|    | 249 |            |            | 1650       |            |            | 1680        |
|    | 250 | AGTTTTTCGT | CCACTGAGCG | TCAGACCCCG | TAGAAAAGAT | CAAAGGATCT | TCTTGAGATC  |
|    | 251 |            |            | 1710       |            |            | 1740        |
|    | 252 | CTTTTTTTCT | GCGCGTAATC | TGCTGCTTGC | AAACAAAAAA | ACCACCGCTA | CCAGCGGTGG  |
| 20 | 253 |            |            | 1770       |            |            | 1800        |
|    | 254 | TTTGTTTGCC | GGATCAAGAG | CTACCAACTC | TTTTTCCGAA | GGTAACTGGC | TTCAGCAGAG  |
|    | 255 |            |            | 1830       |            |            | 1860        |
|    | 256 | CGCAGATACC | AAATACTGTC | CTTCTAGTGT | AGCCGTAGTT | AGGCCACCAC | TTCAAGAACT  |
|    | 257 |            |            | 1890       |            |            | 1920        |
|    | 258 | CTGTAGCACC | GCCTACATAC | CTCGCTCTGC | TAATCCTGTT | ACCAGTGGCT | GCTGCCAGTG  |
| 25 | 259 |            |            | 1950       |            |            | 1980        |
|    | 260 | GCGATAAGTC | GTGTCTTACC | GGGTTGGACT | CAAGACGATA | GTTACCGGAT | AAGGCGCAGC  |
|    | 261 |            |            | 2010       |            |            | 2040        |
|    | 262 | GGTCGGGCTG | AACGGGGGGT | TCGTGCACAC | AGCCAGCTT  | GGAGCGAACG | ACCTACACCG  |
|    | 263 |            |            | 2070       |            |            | 2100        |
|    | 264 | AACTGAGATA | CCTACAGCGT | GAGCTATGAG | AAAGCGCCAC | GCTTCCCGAA | GGGAGAAAGG  |
| 30 | 265 |            |            | 2130       |            |            | 2160        |
|    | 266 | CGGACAGGTA | TCCGGTAAGC | GGCAGGGTCG | GAACAGGAGA | GCGCACGAGG | GAGCTTCCAG  |
|    | 267 |            |            | 2190       |            |            | 2220        |
|    | 268 | GGGGAAACGC | CTGGTATCTT | TATAGTCCTG | TCGGGTTTCG | CCACCTCTGA | CTTGAGCGTC  |
|    | 269 |            |            | 2250       |            |            | 2280        |
| 35 | 270 | GATTTTTGTG | ATGCTCGTCA | GGGGGGCGGA | GCCTATGGAA | AAACGCCAGC | AACGCGGCCT  |
|    | 271 |            |            | 2310       |            |            | 2340        |
|    | 272 | TTTTACGGTT | CCTGGCCTTT | TGCTGGCCTT | TTGCTCACAT | GTTCTTTTCT | GCGTTATCCC  |
|    | 273 |            |            | 2370       |            |            | 2400        |
|    | 274 | CTGATTCTGT | GGATAACCGT | ATTACCGCCT | TTGAGTGAGC | TGATACCGCT | CGCCGCAGCC  |
|    | 275 |            |            | 2430       |            |            | 2460        |
| 40 | 276 | GAACGACCGA | GCGCAGCGAG | TCAGTGAGCG | AGGAAGCGGA | AGAGCGCCTG | ATGCGGTATT  |
|    | 277 |            |            | 2490       |            |            | 2520        |
|    | 278 | TTCTCCTTAC | GCATCTGTGC | GGTATTTTAC | ACCGCATAAA | TTCCGACACC | ATCGAATGGT  |
|    | 279 |            |            | 2550       |            |            | 2580        |
|    | 280 | GCAAAACCTT | TCGCGGTATG | GCATGATAGC | GCCCGGAAGA | GAGTCAATTC | AGGGTGGTGA  |
|    | 281 |            |            | 2610       |            |            | 2640        |
| 45 | 282 | ATGTGAAACC | AGTAACGTTA | TACGATGTCG | CAGAGTATGC | CGGTGTCTCT | TATCAGACCG  |
|    | 283 |            |            | 2670       |            |            | 2700        |
|    | 284 | TTTCCCGCGT | GGTGAACCAG | GCCAGCCACG | TTTCTGCGAA | AACGCGGGAA | AAAGTGGAAG  |
|    | 285 |            |            | 2730       |            |            | 2760        |
|    | 286 | CGGCGATGGC | GGAGCTGAAT | TACATTCCCA | ACCGCGTGGC | ACAACAAC   | GCGGGCAAAC  |
|    | 287 |            |            | 2790       |            |            | 2820        |
| 50 | 288 | AGTCGTTGCT | GATTGGCGTT | GCCACCTCCA | GTCTGGCCCT | GCACGCGCCG | TCGCAAATTG  |
|    | 289 |            |            | 2850       |            |            | 2880        |
|    | 290 | TCGCGGCGAT | TAAATCTCGC | GCCGATCAAC | TGGGTGCCAG | CGTGGTGGTG | TCGATGGTAG  |

55

291 2910 2940  
 292 AACGAAGCGG CGTCGAAGCC TGTAAGCGG CGGTGCACAA TCTTCTCGCG CAACGCGTCA  
 293 2970 3000  
 5 294 GTGGGCTGAT CATTAACTAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCCT  
 295 3030 3060  
 296 GCAC TAATGT TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC AACAGTATTA  
 297 3090 3120  
 298 TTTTCTCCCA TGAAGACGGT ACGCGACTGG GCGTGGAGCA TCTGGTCGCA TTGGGTCACC  
 299 3150 3180  
 10 300 AGCAAATCGC GCTGTTAGCG GGCCCATTA GTTCTGTCTC GGCGCGTCTG CGTCTGGCTG  
 301 3210 3240  
 302 GCTGGCATAA ATATCTCACT CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT  
 303 3270 3300  
 304 GGAGTGCCAT GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA  
 305 3330 3360  
 15 306 CTGCGATGCT GGTGCGCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC ATTACCGAGT  
 307 3390 3420  
 308 CCGGGCTGCG CGTTGGTGCG GATATCTCGG TAGTGGGATA CGACGATACC GAAGACAGCT  
 309 3450 3480  
 20 310 CATGTTATAT CCCGCCGTTA ACCACCATCA AACAGGATTT TCGCCTGCTG GGGCAAACCA  
 311 3510 3540  
 312 GCGTGGACCG CTTGCTGCAA CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC  
 313 3570 3600  
 314 CCGTCTCACT GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC  
 315 3630 3660  
 25 316 GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGGC  
 317 3690 3720  
 318 AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCATCATT AGGCACCCCA GGCTTTACAC  
 319 3750 3780  
 320 TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAATT TCACACAGGA  
 321 3810 3840  
 30 322 AACAGCTATG ACCATGATTA CGGATTCATG GGCCGTCGTT TTACAACGTC GTGACTGGGA  
 323 3870 3900  
 324 AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTTCG CCAGCTGGCG  
 325 3930 3960  
 326 TAATAGCGAA GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC TGAATGGCGA  
 327 3990 4020  
 35 328 ATGGCGCTTT GCCTGGTTTC CGGCACCAGA AGCGGTGCCG GAAAGCTGGC TGGAGTGCGA  
 329 4050 4080  
 330 TCTTCCTGAG GCCGATACTG TCGTCGTCCC CTCAAACCTGG CAGATGCACG GTTACGATGC  
 331 4110 4140  
 332 GCCCATCTAC ACCAACGTAA CCTATCCCAT TACGGTCAAT CCGCCGTTTG TTCCACGGA  
 333 4170 4200  
 40 334 GAATCCGACG GGTTGTTACT CGCTCACATT TAATGTTGAT GAAAGCTGGC TACAGGAAGG  
 335 4230 4260  
 336 CCAGACGCGA ATTATTTTTG ATGGCGTTGG AATTACGTTA TCGACTGCAC GGTGCACCAA  
 337 4290 4320  
 338 TGCTTCTGGC GTCAGGCAGC CATCGGAAGC TGTGGTATGG CTGTGCAGGT CGTAAATCAC  
 339 4350 4380  
 45 340 TGCATAATTC GTGTCGCTCA AGGCGCACTC CCGTTCTGGA TAATGTTTTT TGCGCCGACA  
 341 4410 4440  
 342 TCATAACGGT TCTGGCAAAT GGGAATTGGG AAATTAATAC GACTCACTAT ATGGAATTGT  
 343 4470 4500  
 344 GAGCGGATAA CAATTCCTAG AAATAATTTT GTTTAACTTT AAGAAGGAGA TATACAT  
 345 4530  
 50 346  
 347  
 348

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

349 (A) LENGTH: 672 nucleic acids  
 350 (B) TYPE: nucleic acid  
 351 (C). STRANDEDNESS: double strand  
 5 352 (D) TOPOLOGY: linear  
 353 (ii) MOLECULE TYPE: other nucleic acid, genomic DNA  
 354 (vi) ORIGINAL SOURCE:  
 355 (A) ORGANISM: synthesized, HTLV-I  
 356 (x) PUBLICATION INFORMATION:  
 357 (A) AUTHORS: Nobuyuki FUJII et al.  
 10 358 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 359 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 360 (K) RELEVANT RESIDUES IN SEQ ID NO:6: FROM 1 to 672  
 361  
 362 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:  
 363  
 15 364 ATGGCGTGGG AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC  
 365 30 60  
 366 CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCAACCAAAA GGTAGAGGTT  
 367 90 120  
 368 ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT  
 369 150 180  
 20 370 GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGATC GGAATTCATG  
 371 210 240  
 372 GGCCAAATCT TTTCCCGTAG CGCTAGCCCT ATTCCGCGGC CGCCCCGGGG GCTGGCCGCT  
 373 270 300  
 374 CATCACTGGC TTAACCTCCT CCAGGCGGCA TATCGCCTAG AACCCGGTCC CTCCAGTTAC  
 375 330 360  
 25 376 GATTTCCACC AGTTAAAAAA ATTTCTTAAA ATAGCTTTAG AAACACCGGT CTGGATCTGC  
 377 390 420  
 378 CCCATTAACT ACTCCCTCCT AGCCAGCCTA CTCCCAAAAG GATACCCCGG CCGGGTGAAT  
 379 450 480  
 380 GAAATTTTAC ACATACTCAT CCAAACCCAA GCCCAGATCC CGTCCCCGCC CGCGCCGCCG  
 381 510 540  
 30 382 CCGCCGTCAT CCTCCACCCA CGACCCCGG GATTCTGACC CACAAATCCC CCCTCCCTAT  
 383 570 600  
 384 GTTGAGCCTA CAGCCCCCA AGTCCTTTAA GGATCCGGGC CCTCTAGATG CGGCCGCATG  
 385 630 660  
 386 CATGGTACCT AA  
 387  
 35 388 (2) INFORMATION FOR SEQ ID NO: 7:  
 389 (i) SEQUENCE CHARACTERISTICS:  
 390 (A) LENGTH: 224 amino acids  
 391 (B) TYPE: amino acid  
 392 (D) TOPOLOGY: linear  
 40 393 (ii) MOLECULE TYPE: protein  
 394 (vi) ORIGINAL SOURCE:  
 395 (A) ORGANISM: recombinant  
 396 (x) PUBLICATION INFORMATION:  
 397 (A) AUTHORS: Nobuyuki FUJII et al.  
 398 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 45 399 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 400 (K) RELEVANT RESIDUES IN SEQ ID NO:7: FROM 1 to 224  
 401  
 402 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:  
 403  
 404 Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser  
 50 405 1 20  
 406 Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val

55

407 21 40  
 408 Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser  
 409 41 60  
 5 410 Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Met  
 411 61 80  
 412 Gly Gln Ile Phe Ser Arg Ser Ala Ser Pro Ile Pro Arg Pro Pro Arg Gly Leu Ala Ala  
 413 81 100  
 414 His His Trp Leu Asn Phe Leu Gln Ala Ala Tyr Arg Leu Glu Pro Gly Pro Ser Ser Tyr  
 415 101 120  
 10 416 Asp Phe His Gln Leu Lys Lys Phe Leu Lys Ile Ala Leu Glu Thr Pro Val Trp Ile Cys  
 417 121 140  
 418 Pro Ile Asn Tyr Ser Leu Leu Ala Ser Leu Leu Pro Lys Gly Tyr Pro Gly Arg Val Asn  
 419 141 160  
 420 Glu Ile Leu His Ile Leu Ile Gln Thr Gln Ala Gln Ile Pro Ser Arg Pro Ala Pro Pro  
 15 421 161 180  
 422 Pro Pro Ser Ser Ser Thr His Asp Pro Pro Asp Ser Asp Pro Gln Ile Pro Pro Pro Tyr  
 423 181 200  
 424 Val Glu Pro Thr Ala Pro Gln Val Leu \*\*\* Gly Ser Gly Pro Ser Arg Cys Gly Arg Met  
 425 201 220  
 426 His Gly Thr \*\*\*  
 20 427 221  
 428  
 429 (2) INFORMATION FOR SEQ ID NO: 8:  
 430 (i) SEQUENCE CHARACTERISTICS:  
 431 (A) LENGTH: 690 nucleic acids  
 432 (B) TYPE: nucleic acid  
 25 433 (C) STRANDEDNESS: double strand  
 434 (D) TOPOLOGY: linear  
 435 (ii) MOLECULE TYPE: other nucleic acid, genomic DNA  
 436 (vi) ORIGINAL SOURCE:  
 437 (A) ORGANISM: synthesized, HTLV-II  
 438 (x) PUBLICATION INFORMATION:  
 30 439 (A) AUTHORS: Nobuyuki FUJII et al.  
 440 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 441 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 442 (K) RELEVANT RESIDUES IN SEQ ID NO:8: FROM 1 to 690  
 443  
 35 444 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:  
 445  
 446 ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC  
 447 30 60  
 448 CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCCAACCAAA GGTAGAGGTT  
 449 90 120  
 40 450 ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT  
 451 150 180  
 452 GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGGATC GGAATTCATG  
 453 210 240  
 454 GGACAAATCC ACGGGCTTTC CCCAACTCCA ATACCCAAAG CCCCAGGGG GCTATCAACC  
 455 270 300  
 45 456 CACCACTGGC TTAACCTTCT CCAGGCTGCT TACCGCTTGC AGCCTAGGCC CTCCGATTTC  
 457 330 360  
 458 GACTTCCAGC AGCTACGACG CTTTCTAAAA CTAGCCCTTA AAACGCCCCAT TTGGCTAAAT  
 459 390 420  
 460 CCTATTGACT ACTCGCTTTT AGCTAGCCTT ATCCCCAAGG GATATCCAGG AAGGGTGGTA  
 461 450 480  
 50 462 GAGATTATAA ATATCCTTGT CAAAAATCAA GTCTCCCCTA GCGCCCCCGC CGCCCCAGTT  
 463 510 540  
 464 CCGACACCTA TCTGCCCTAC TACTACTCCT CCGCCACCTC CCCCCCTTC CCCGGAGGCC

55

465 570 600  
 466 CATGTTCCCC CCCCTTACGT GGAACCCACC ACCACGCAAT GCTTCTAAGG ATCCGGGCCCC  
 467 630 660  
 5 468 TCTAGATGCG GCCGCATGCA TGGTACCTAA  
 469 690  
 470  
 471 (2) INFORMATION FOR SEQ ID NO: 9:  
 472 (i) SEQUENCE CHARACTERISTICS:  
 473 (A) LENGTH: 230 amino acids  
 10 474 (B) TYPE: amino acid  
 475 (D) TOPOLOGY: linear  
 476 (ii) MOLECULE TYPE: protein  
 477 (vi) ORIGINAL SOURCE:  
 478 (A) ORGANISM: recombinant  
 479 (x) PUBLICATION INFORMATION:  
 15 480 (A) AUTHORS: Nobuyuki FUJII et al.  
 481 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 482 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 483 (K) RELEVANT RESIDUES IN SEQ ID NO:9: FROM 1 to 230  
 484  
 20 485 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:  
 486  
 487 Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser  
 488 1 20  
 489 Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val  
 490 21 40  
 25 491 Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser  
 492 41 60  
 493 Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Met  
 494 61 80  
 495 Gly Gln Ile His Gly Leu Ser Pro Thr Pro Ile Pro Lys Ala Pro Arg Gly Leu Ser Thr  
 496 81 100  
 30 497 His His Trp Leu Asn Phe Leu Gln Ala Ala Tyr Arg Leu Gln Pro Arg Pro Ser Asp Phe  
 498 101 120  
 499 Asp Phe Gln Gln Leu Arg Arg Phe Leu Lys Leu Ala Leu Lys Thr Pro Ile Trp Leu Asn  
 500 121 140  
 501 Pro Ile Asp Tyr Ser Leu Leu Ala Ser Leu Ile Pro Lys Gly Tyr Pro Gly Arg Val Val  
 502 141 160  
 35 503 Glu Ile Ile Asn Ile Leu Val Lys Asn Gln Val Ser Pro Ser Ala Pro Ala Ala Pro Val  
 504 161 180  
 505 Pro Thr Pro Ile Cys Pro Thr Thr Thr Pro Pro Pro Pro Pro Pro Ser Pro Glu Ala  
 506 181 200  
 507 His Val Pro Pro Pro Tyr Val Glu Pro Thr Thr Thr Gln Cys Phe \*\*\* Gly Ser Gly Pro  
 40 508 201 220  
 509 Ser Arg Cys Gly Arg Met His Gly Thr \*\*\*  
 510 221  
 511  
 512 (2) INFORMATION FOR SEQ ID NO: 10:  
 513 (i) SEQUENCE CHARACTERISTICS:  
 45 514 (A) LENGTH: 810 nucleic acids  
 515 (B) TYPE: nucleic acid  
 516 (C) STRANDEDNESS: double strand  
 517 (D) TOPOLOGY: linear  
 518 (ii) MOLECULE TYPE: other nucleic acid, genomic DNA  
 519 (vi) ORIGINAL SOURCE:  
 50 520 (A) ORGANISM: synthesized, HTLV-I  
 521 (x) PUBLICATION INFORMATION:  
 522 (A) AUTHORS: Nobuyuki FUJII et al.

55

(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 (K) RELEVANT RESIDUES IN SEQ ID NO:10: FROM 1 to 810

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC  
 30 60  
 CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCCAACCAAA GGTAGAGGTT  
 90 120  
 ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT  
 150 180  
 GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGATC GGAATTCGCA  
 210 240  
 GTACCGGTGG CGGTCTGGCT TGTCTCCGCC CTGGCCATGG GAGCCGGAGT GGCTGGCAGG  
 270 300  
 ATTACCGGCT CCATGTCCCT CGCCTCAGGA AAGAGCCTCC TACATGAGGT GGACAAAGAT  
 330 360  
 ATTTCCCAAT TAAC'TCAAGC AATAGTCAAA AACCACAAAA ATCTGCTCAA AATTGCACAG  
 390 420  
 TATGCTGCCC AGAACAGACG AGGCCTTGAT CTCCTGTTCT GGGAGCAAGG AGGATTATGC  
 450 480  
 AAAGCATTAC AAGAACAGTG CTGTTTTCTA AATATTACTA ATTCCCATGT CTCAATACTA  
 510 540  
 CAAGAGAGAC CCCCCCTTGA AAATCGAGTC CTGACTGGCT GGGGCCTTAA CTGGGACCTT  
 570 600  
 GGCCTCTCAC AGTGGGCTCG AGAAGCCTTA CAAACTGGAA TCACCCTTGT CGCGCTACTC  
 630 660  
 CTTCTTGTTA TCCTTGCAGG ACCATGCATC CTCCGTCAGC TACGACACCT CCCCTCGCGC  
 690 720  
 GTCAGATACC CCCATTACTC TCTTATAAAC CCTGAGTCAT CCCTGTAAGG ATCCGGGCCC  
 750 780  
 TCTAGATGCG GCCGCATGCA TGGTACCTAA  
 810

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 270 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

(A) ORGANISM: recombinant

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Nobuyuki FUJII et al.

(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN

(K) RELEVANT RESIDUES IN SEQ ID NO:11: FROM 1 to 270

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser  
 1 20  
 Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val  
 21 40  
 Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser  
 41 60  
 Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Ala



581 61 80  
582 Val Pro Val Ala Val Trp Leu Val Ser Ala Leu Ala Met Gly Ala Gly Val Ala Gly Arg  
583 81 100  
5 584 Ile Thr Gly Ser Met Ser Leu Ala Ser Gly Lys Ser Leu Leu His Glu Val Asp Lys Asp  
585 101 120  
586 Ile Ser Gln Leu Thr Gln Ala Ile Val Lys Asn His Lys Asn Leu Leu Lys Ile Ala Gln  
587 121 140  
588 Tyr Ala Ala Gln Asn Arg Arg Gly Leu Asp Leu Leu Phe Trp Glu Gln Gly Gly Leu Cys  
589 141 160  
10 590 Lys Ala Leu Gln Glu Gln Cys Cys Phe Leu Asn Ile Thr Asn Ser His Val Ser Ile Leu  
591 161 180  
592 Gln Glu Arg Pro Pro Leu Glu Asn Arg Val Leu Thr Gly Trp Gly Leu Asn Trp Asp Leu  
593 181 200  
594 Gly Leu Ser Gln Trp Ala Arg Glu Ala Leu Gln Thr Gly Ile Thr Leu Val Ala Leu Leu  
595 201 220  
15 596 Leu Leu Val Ile Leu Ala Gly Pro Cys Ile Leu Arg Gln Leu Arg His Leu Pro Ser Arg  
597 221 240  
598 Val Arg Tyr Pro His Tyr Ser Leu Ile Asn Pro Glu Ser Ser Leu \*\*\* Gly Ser Gly Pro  
599 241 260  
600 Ser Arg Cys Gly Arg Met His Gly Thr \*\*\*  
601 261  
20 602  
603 (2) INFORMATION FOR SEQ ID NO: 12:  
604 (i) SEQUENCE CHARACTERISTICS:  
605 (A) LENGTH: 816 nucleic acids  
606 (B) TYPE: nucleic acid  
25 607 (C) STRANDEDNESS: double strand  
608 (D) TOPOLOGY: linear  
609 (ii) MOLECULE TYPE: other nucleic acid, genomic DNA  
610 (vi) ORIGINAL SOURCE:  
611 (A) ORGANISM: synthesized, HTLV-II  
612 (x) PUBLICATION INFORMATION:  
30 613 (A) AUTHORS: Nobuyuki FUJII et al.  
614 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
615 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
616 (K) RELEVANT RESIDUES IN SEQ ID NO:12: FROM 1 to 816  
617  
618 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:  
35 619  
620 ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC  
621 30 60  
622 CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCAACCAAAA GGTAGAGGTT  
623 90 120  
624 ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT  
40 625 150 180  
626 GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGGATC GGAATTCGCC  
627 210 240  
628 GTTCCAATAG CAGTGTGGCT TGTCTCCGCC CTAGCGGCCG GAACAGGTAT CGCTGGTGGA  
629 270 300  
630 GTAACAGGCT CCCTATCTCT GGCTTCCAGT AAAAGCCTTC TCCTCGAGGT TGACAAAGAC  
45 631 330 360  
632 ATCTCCCACC TTACCCAGGC CATAGTCAAA AATCATCAAA ACATCCTCCG GGTTCACACAG  
633 390 420  
634 TATGCAGCCC AAAATAGACG AGGATTAGAC CTCCTATTCT GGAACAAGG GGGTTTGTGC  
635 450 480  
50 636 AAGGCCATAC AGGAGCAATG TTGCTTCCTC AACATCAGTA AACTCATGT ATCCGTCCTC  
637 510 540  
638 CAGGAACGGC CCCCTCTTGA AAAACGTGTC ATCACC GGCT GGGGACTAAA CTGGGATCTT

55

639 570 600  
 640 GGACTGTCCC AATGGGCACG AGAAGCCCTC CAGACAGGCA TAACCATTCT CGCTCTACTC  
 641 630 660  
 5 642 CTCCTCGTCA TATTGTTTGG CCCCTGTATC CTCCGCCAAA TCCAGGCCCT TCCACAGCGG  
 643 690 720  
 644 TTACAAAACC GACATAACCA GTATTCCCTT ATCAACCCAG AAACCATGCT ATAAGGATCC  
 645 750 780  
 646 GGGCCCTCTA GATGCGGCCG CATGCATGGT ACCTAA  
 647 810  
 10 648  
 649 (2) INFORMATION FOR SEQ ID NO: 13:  
 650 (i) SEQUENCE CHARACTERISTICS:  
 651 (A) LENGTH: 272 amino acids  
 652 (B) TYPE: amino acid  
 653 (D) TOPOLOGY: linear  
 15 654 (ii) MOLECULE TYPE: protein  
 655 (vi) ORIGINAL SOURCE:  
 656 (A) ORGANISM: recombinant  
 657 (x) PUBLICATION INFORMATION:  
 658 (A) AUTHORS: Nobuyuki FUJII et al.  
 659 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 20 660 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 661 (K) RELEVANT RESIDUES IN SEQ ID NO:13: FROM 1 to 272  
 662  
 663 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:  
 664  
 25 665 Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser  
 666 1 20  
 667 Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val  
 668 21 40  
 669 Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser  
 670 41 60  
 30 671 Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Ala  
 672 61 80  
 673 Val Pro Ile Ala Val Trp Leu Val Ser Ala Leu Ala Ala Gly Thr Gly Ile Ala Gly Gly  
 674 81 100  
 675 Val Thr Gly Ser Leu Ser Leu Ala Ser Ser Lys Ser Leu Leu Leu Glu Val Asp Lys Asp  
 676 101 120  
 35 677 Ile Ser His Leu Thr Gln Ala Ile Val Lys Asn His Gln Asn Ile Leu Arg Val Ala Gln  
 678 121 140  
 679 Tyr Ala Ala Gln Asn Arg Arg Gly Leu Asp Leu Leu Phe Trp Glu Gln Gly Gly Leu Cys  
 680 141 160  
 681 Lys Ala Ile Gln Glu Gln Cys Cys Phe Leu Asn Ile Ser Asn Thr His Val Ser Val Leu  
 682 161 180  
 40 683 Gln Glu Arg Pro Pro Leu Glu Lys Arg Val Ile Thr Gly Trp Gly Leu Asn Trp Asp Leu  
 684 181 200  
 685 Gly Leu Ser Gln Trp Ala Arg Glu Ala Leu Gln Thr Gly Ile Thr Ile Leu Ala Leu Leu  
 686 201 220  
 687 Leu Leu Val Ile Leu Phe Gly Pro Cys Ile Leu Arg Gln Ile Gln Ala Leu Pro Gln Arg  
 688 221 240  
 45 689 Leu Gln Asn Arg His Asn Gln Tyr Ser Leu Ile Asn Pro Glu Thr Met Leu \*\*\* Gly Ser  
 690 241 260  
 691 Gly Pro Ser Arg Cys Gly Arg Met His Gly Thr \*\*\*  
 692 261  
 693  
 50 694 (2) INFORMATION FOR SEQ ID NO: 14:  
 695 (i) SEQUENCE CHARACTERISTICS:  
 696 (A) LENGTH: 1119 nucleic acids

697 (B) TYPE: nucleic acid  
 698 (C) STRANDEDNESS: double strand  
 699 (D) TOPOLOGY: linear  
 5 700 (ii) MOLECULE TYPE: other nucleic acid, genomic DNA  
 701 (vi) ORIGINAL SOURCE:  
 702 (A) ORGANISM: plasmid, Tp  
 703 (B) STRAIN: Nichols  
 704 (x) PUBLICATION INFORMATION:  
 705 (A) AUTHORS: Nobuyuki FUJII et al.  
 10 706 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 707 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 708 (K) RELEVANT RESIDUES IN SEQ ID NO:14: FROM 1 to 1119  
 709  
 710 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:  
 15 711  
 712 ATGTCCCCTA TACTAGGTTA TTGGAAAATT AAGGGCCTTG TGCAACCCAC TCGACTTCTT  
 713 30 60  
 714 TTGGAATATC TTGAAGAAAA ATATGAAGAG CATTGTATG AGCGCGATGA AGGTGATAAA  
 715 90 120  
 716 TGGCGAAACA AAAAGTTTGA ATTGGGTTTG GAGTTTCCCA ATCTTCCTTA TTATATTGAT  
 20 717 150 180  
 718 GGTGATGTTA AATTAACACA GTCTATGGCC ATCATACGTT ATATAGCTGA CAAGCACAAC  
 719 210 240  
 720 ATGTTGGGTG GTTGTCCAAA AGAGCGTGCA GAGATTTCAA TGCTTGAAGG AGCGGTTTTG  
 721 270 300  
 25 722 GATATTAGAT ACGGTGTTTC GAGAATTGCA TATAGTAAAG ACTTTGAAAC TCTCAAAGTT  
 723 330 360  
 724 GATTTTCTTA GCAAGCTACC TGAAATGCTG AAAATGTTTCG AAGATCGTTT ATGTCATAAA  
 725 390 420  
 726 ACATATTTAA ATGGTGATCA TGTAACCCAT CCTGACTTCA TGTGTATGA CGCTCTTGAT  
 727 450 480  
 30 728 GTTGTTTTAT ACATGGACCC AATGTGCCTG GATGCGTTCC CAAAATTAGT TTGTTTTAAA  
 729 510 540  
 730 AAACGTATTG AAGCTATCCC ACAAATTGAT AAGTACTTGA AATCCAGCAA GTATATAGCA  
 731 570 600  
 732 TGGCCTTTGC AGGGCTGGCA AGCCACGTTT GGTGGTGGCG ACCATCCTCC AAAATCGGAT  
 733 630 660  
 35 734 CTGGTTCCGC GTGGATCGGA ATTCTGTTCA TTTAGTTCTA TCCCGAATGG CACGTACCGG  
 735 690 720  
 736 GCGACGTATC AGGATTTTGA TGAGAATGGT TGGAAGGACT TTCTCGAGGT TACTTTTGAT  
 737 750 780  
 738 GGTGGCAAGA TGGTGCAGGT GGTTTACGAT TATCAGCATA AAGAAGGGCG GTTTAAGTCC  
 739 810 840  
 40 740 CAGGACGCTG ACTACCATCG GGTCATGTAT GCATCCTCGG GCATAGGTCC TGAAAAGGCC  
 741 870 900  
 742 TTCAGAGAGC TCGCCGATGC TTTGCTTGAA AAGGGTAATC CCGAGATGGT GGATGTGGTC  
 743 930 960  
 744 ACCGGTGCAA CTGTTTCTTC CCAGAGTTTC AGGAGTTTGG GTCGTGCGCT TCTGCAGAGT  
 45 745 990 1020  
 746 GCGCGGCGCG GCGAGAAGGA AGCCATTATT AGCAGGTAGG AATTCGTCGA CCTCGAGGGA  
 747 1050 1080  
 748 TCCGGGCCCT CTAGATGCGG CCGCATGCAT GGTACCTAA  
 749 1110  
 750  
 50 751 (2) INFORMATION FOR SEQ ID NO: 15:  
 752 (i) SEQUENCE CHARACTERISTICS:  
 753 (A) LENGTH: 373 amino acids  
 754 (B) TYPE: amino acid

755 (D) TOPOLOGY: linear  
 756 (ii) MOLECULE TYPE: protein  
 757 (vi) ORIGINAL SOURCE:  
 5 758 (A) ORGANISM: recombinant  
 759 (x) PUBLICATION INFORMATION:  
 760 (A) AUTHORS: Nobuyuki FUJII et al.  
 761 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 762 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 763 (K) RELEVANT RESIDUES IN SEQ ID NO:15: FROM 1 to 373  
 10 764  
 765 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:  
 766  
 767 Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro Thr Arg Leu Leu  
 768 1 20  
 15 769 Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu Tyr Glu Arg Asp Glu Gly Asp Lys  
 770 21 40  
 771 Trp Arg Asn Lys Lys Phe Glu Leu Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp  
 772 41 60  
 773 Gly Asp Val Lys Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
 774 61 80  
 20 775 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu Gly Ala Val Leu  
 776 81 100  
 777 Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser Lys Asp Phe Glu Thr Leu Lys Val  
 778 101 120  
 779 Asp Phe Leu Ser Lys Leu Pro Glu Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys  
 780 121 140  
 25 781 Thr Tyr Leu Asn Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
 782 141 160  
 783 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu Val Cys Phe Lys  
 784 161 180  
 785 Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr Leu Lys Ser Ser Lys Tyr Ile Ala  
 786 181 200  
 30 787 Trp Pro Leu Gln Gly Trp Gln Ala Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp  
 788 201 220  
 789 Leu Val Pro Arg Gly Ser Glu Phe Cys Ser Phe Ser Ser Ile Pro Asn Gly Thr Tyr Arg  
 790 221 240  
 791 Ala Thr Tyr Gln Asp Phe Asp Glu Asn Gly Trp Lys Asp Phe Leu Glu Val Thr Phe Asp  
 792 241 260  
 35 793 Gly Gly Lys Met Val Gln Val Val Tyr Asp Tyr Gln His Lys Glu Gly Arg Phe Lys Ser  
 794 261 280  
 795 Gln Asp Ala Asp Tyr His Arg Val Met Tyr Ala Ser Ser Gly Ile Gly Pro Glu Lys Ala  
 796 281 300  
 797 Phe Arg Glu Leu Ala Asp Ala Leu Leu Glu Lys Gly Asn Pro Glu Met Val Asp Val Val  
 798 301 320  
 40 799 Thr Gly Ala Thr Val Ser Ser Gln Ser Phe Arg Arg Leu Gly Arg Ala Leu Leu Gln Ser  
 800 321 340  
 801 Ala Arg Arg Gly Glu Lys Glu Ala Ile Ile Ser Arg \*\*\* Glu Phe Val Asp Leu Glu Gly  
 802 341 360  
 803 Ser Gly Pro Ser Arg Cys Gly Arg Met His Gly Thr \*\*\*  
 804 361  
 45 805  
 806 (2) INFORMATION FOR SEQ ID NO: 16:  
 807 (i) SEQUENCE CHARACTERISTICS:  
 808 (A) LENGTH: 858 nucleic acids  
 809 (B) TYPE: nucleic acid  
 50 810 (C) STRANDEDNESS: double strand  
 811 (D) TOPOLOGY: linear  
 812 (ii) MOLECULE TYPE: genomic DNA

55

813 (vi) ORIGINAL SOURCE:  
814 (A) ORGANISM: *E. coli*, Tp  
815 (B). STRAIN: DH15 $\alpha$ , Nichols  
5 816 (x) PUBLICATION INFORMATION:  
817 (A) AUTHORS: Nobuyuki FUJII et al.  
818 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
819 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
820 (K) RELEVANT RESIDUES IN SEQ ID NO:16: FROM 1 to 858  
10 821  
822 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:  
823  
824 ATGTTACACC AACAAACGAAA CCAACACGCC AGGCTTATTC CTGTGGAGTT ATATATGAGC  
825 30 60  
826 GATAAAATTA TTCACCTGAC TGACGACAGT TTTGACACGG ATGTACTCAA AGCGGACGGG  
15 827 90 120  
828 GCGATCCTCG TCGATTTCTG GGCAGAGTGG TGCGGTCCGT GCAAAATGAT CGCCCCGATT  
829 150 180  
830 CTGGATGAAA TCGCTGACGA ATATCAGGGC AAAGTACCG TTGCAAAACT GAACATCGAT  
831 210 240  
832 CAAAACCCCTG GCACTGCGCC GAAATATGGC ATCCGTGGTA TCCCGACTCT GCTGCTGTTC  
20 833 270 300  
834 AAAAACGGTG AAGTGGCGGC AACCAAAGTG GGTGCACTGT CTAAAGGTCA GTTGAAAGAG  
835 330 360  
836 TTCCTCGACG CTAACCTGGC GGAGCTCGGT GGTCTTCTC TGGTTCCGCG TGGATCGGAA  
837 390 420  
838 TTCTGTTCAT TTAGTTCTAT CCCGAATGGC ACGTACCGG CGACGTATCA GGATTTTGAT  
25 839 450 480  
840 GAGAATGGTT GGAAGGACTT TCTCGAGGTT ACTTTTGATG GTGGCAAGAT GGTGCAGGTG  
841 510 540  
842 GTTTACGATT ATCAGCATAA AGAAGGGCGG TTTAAGTCCC AGGACGCTGA CTACCATCGG  
843 570 600  
30 844 GTCATGTATG CATCCTCGGG CATAGGTCCT GAAAAGGCCT TCAGAGAGCT CGCCGATGCT  
845 630 660  
846 TTGCTTGAAA AGGGTAATCC CGAGATGGTG GATGTGGTCA CCGGTGCAAC TGTTTCTTCC  
847 690 720  
848 CAGAGTTTCA GGAGGTTGGG TCGTGCCTT CTGCAGAGTG CGCGGCGCGG CGAGAAGGAA  
849 750 780  
35 850 GCCATTATTA GCAGGTAGGA ATTCGTCGAC CTCGAGGGAT CCGGGCCCTC TAGATGCGGC  
851 810 840  
852 CGCATGCATG GTACCTAA  
853  
854  
40 855 (2) INFORMATION FOR SEQ ID NO: 17:  
856 (i) SEQUENCE CHARACTERISTICS:  
857 (A) LENGTH: 286 amino acids  
858 (B) TYPE: amino acid  
859 (D) TOPOLOGY: linear  
860 (ii) MOLECULE TYPE: protein  
45 861 (vi) ORIGINAL SOURCE:  
862 (A) ORGANISM: recombinant  
863 (x) PUBLICATION INFORMATION:  
864 (A) AUTHORS: Nobuyuki FUJII et al.  
865 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
50 866 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
867 (K) RELEVANT RESIDUES IN SEQ ID NO:17: FROM 1 to 286  
868  
869 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:  
870

55

871 Met Leu His Gln Gln Arg Asn Gln His Ala Arg Leu Ile Pro Val Glu Leu Tyr Met Ser  
 872 1 20  
 873 Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly  
 5 874 21 40  
 875 Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile  
 876 41 60  
 877 Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp  
 878 61 80  
 879 Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe  
 10 880 81 100  
 881 Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu  
 882 101 120  
 883 Phe Leu Asp Ala Asn Leu Ala Glu Leu Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu  
 884 121 140  
 885 Phe Cys Ser Phe Ser Ser Ile Pro Asn Gly Thr Tyr Arg Ala Thr Tyr Gln Asp Phe Asp  
 15 886 141 160  
 887 Glu Asn Gly Trp Lys Asp Phe Leu Glu Val Thr Phe Asp Gly Gly Lys Met Val Gln Val  
 888 161 180  
 889 Val Tyr Asp Tyr Gln His Lys Glu Gly Arg Phe Lys Ser Gln Asp Ala Asp Tyr His Arg  
 890 181 200  
 891 Val Met Tyr Ala Ser Ser Gly Ile Gly Pro Glu Lys Ala Phe Arg Glu Leu Ala Asp Ala  
 20 892 201 220  
 893 Leu Leu Glu Lys Gly Asn Pro Glu Met Val Asp Val Val Thr Gly Ala Thr Val Ser Ser  
 894 221 240  
 895 Gln Ser Phe Arg Arg Leu Gly Arg Ala Leu Leu Gln Ser Ala Arg Arg Gly Glu Lys Glu  
 896 241 260  
 897 Ala Ile Ile Ser Arg \*\*\* Glu Phe Val Asp Leu Glu Gly Ser Gly Pro Ser Arg Cys Gly  
 25 898 261 280  
 899 Arg Met His Gly Thr \*\*\*  
 900 281  
 901  
 902 (2) INFORMATION FOR SEQ ID NO: 18:  
 903 (i) SEQUENCE CHARACTERISTICS:  
 904 (A) LENGTH: 672 nucleic acids  
 905 (B) TYPE: nucleic acid  
 906 (C) STRANDEDNESS: double strand  
 907 (D) TOPOLOGY: linear  
 908 (ii) MOLECULE TYPE: other nucleic acid, genomic DNA  
 35 909 (vi) ORIGINAL SOURCE:  
 910 (A) ORGANISM: synthesized, Tp  
 911 (B) STRAIN: Nichols  
 912 (x) PUBLICATION INFORMATION:  
 913 (A) AUTHORS: Nobuyuki FUJII et al.  
 914 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 40 915 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 916 (K) RELEVANT RESIDUES IN SEQ ID NO:18: FROM 1 to 672  
 917  
 918 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:  
 919  
 920 ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC  
 45 921 30 60  
 922 CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCAACCAAA GGTAGAGGTT  
 923 90 120  
 924 ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT  
 925 150 180  
 926 GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGATC GGAATTCTGT  
 50 927 210 240  
 928 TCATTTAGTT CTATCCCGAA TGGCACGTAC CGGGCGACGT ATCAGGATTT TGATGAGAAT

55

929 270 300  
 930 GGTGGAAGG ACTTCTCGA GGTTACTTTT GATGGTGGCA AGATGGTGCA GGTGGTTTAC  
 931 330 360  
 5 932 GATTATCAGC ATAAAGAAGG GCGGTTTAAAG TCCCAGGACG CTGACTACCA TCGGGTCATG  
 933 390 420  
 934 TATGCATCCT CGGGCATAGG TCCTGAAAAG GCCTTCAGAG AGCTCGCCGA TGCTTTGCTT  
 935 450 480  
 936 GAAAAGGGTA ATCCCGAGAT GGTGGATGTG GTCACCGGTG CAACTGTTTC TTCCCAGAGT  
 937 510 540  
 10 938 TTCAGGAGGT TGGGTCGTGC GCTTCTGCAG AGTGCGCGGC GCGGCGAGAA GGAAGCCATT  
 939 570 600  
 940 ATTAGCAGGT AGGAATTCGT CGACCTCGAG GGATCCGGGC CCTCTAGATG CGGCCGCATG  
 941 630 660  
 942 CATGGTACCT AA  
 943  
 15 944 (2) INFORMATION FOR SEQ ID NO: 19:  
 945 (i) SEQUENCE CHARACTERISTICS:  
 946 (A) LENGTH: 224 amino acids  
 947 (B) TYPE: amino acid  
 948 (D) TOPOLOGY: linear  
 949 (ii) MOLECULE TYPE: protein  
 20 950 (vi) ORIGINAL SOURCE:  
 951 (A) ORGANISM: recombinant  
 952 (x) PUBLICATION INFORMATION:  
 953 (A) AUTHORS: Nobuyuki FUJII et al.  
 954 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 955 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 25 956 (K) RELEVANT RESIDUES IN SEQ ID NO:19: FROM 1 to 224  
 957  
 958 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:  
 959  
 960 Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser  
 961 1 20  
 30 962 Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val  
 963 21 40  
 964 Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser  
 965 41 60  
 966 Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Cys  
 967 61 80  
 35 968 Ser Phe Ser Ser Ile Pro Asn Gly Thr Tyr Arg Ala Thr Tyr Gln Asp Phe Asp Glu Asn  
 969 81 100  
 970 Gly Trp Lys Asp Phe Leu Glu Val Thr Phe Asp Gly Gly Lys Met Val Gln Val Val Tyr  
 971 101 120  
 972 Asp Tyr Gln His Lys Glu Gly Arg Phe Lys Ser Gln Asp Ala Asp Tyr His Arg Val Met  
 973 121 140  
 40 974 Tyr Ala Ser Ser Gly Ile Gly Pro Glu Lys Ala Phe Arg Glu Leu Ala Asp Ala Leu Leu  
 975 141 160  
 976 Glu Lys Gly Asn Pro Glu Met Val Asp Val Val Thr Gly Ala Thr Val Ser Ser Gln Ser  
 977 161 180  
 978 Phe Arg Arg Leu Gly Arg Ala Leu Leu Gln Ser Ala Arg Arg Gly Glu Lys Glu Ala Ile  
 979 181 200  
 45 980 Ile Ser Arg \*\*\* Glu Phe Val Asp Leu Glu Gly Ser Gly Pro Ser Arg Cys Gly Arg Met  
 981 201 220  
 982 His Gly Thr \*\*\*  
 983 221  
 984  
 50 985 (2) INFORMATION FOR SEQ ID NO: 20:  
 986 (i) SEQUENCE CHARACTERISTICS:

55

987 (A) LENGTH: 1035 nucleic acids  
 988 (B) TYPE: nucleic acid  
 989 (C) STRANDEDNESS: double strand  
 5 990 (D) TOPOLOGY: linear  
 991 (ii) MOLECULE TYPE: other nucleic acid, genomic DNA  
 992 (vi) ORIGINAL SOURCE:  
 993 (A) ORGANISM: synthesized, Tp  
 994 (B) STRAIN: Nichols  
 10 995 (x) PUBLICATION INFORMATION:  
 996 (A) AUTHORS: Nobuyuki FUJII et al.  
 997 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 998 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 999 (K) RELEVANT RESIDUES IN SEQ ID NO:20: FROM 1 to 1035  
 1000  
 1001 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:  
 15 1002  
 1003 ATGAAGATTG GTATTGTAAC TGGTATCCCT GGTGTAGGGA AAAGTACTGT CTTGGCTAAA  
 1004 30 60  
 1005 GTTAAAGAGA TATTGGATAA TCAAGGTATA AATAACAAGA TCATAAATTA TGGAGATTTT  
 1006 90 120  
 20 1007 ATGTTAGCAA CAGCATTAAG ATTAGGCTAT GCTAAAGATA GAGACGAAAT GAGAAAATTA  
 1008 150 180  
 1009 TCTGTAGAAA AGCAGAAGAA ATTGCAGATT GATGCGGCTA AAGGTATAGC TGAAGAGGCA  
 1010 210 240  
 1011 AGAGCAGGTG GAGAAGGATA TCTGTTTCATA GATACGCACG CTGTGATACG TACACCCTCT  
 1012 270 300  
 25 1013 GGATATTTAC CTGGTTTACC GTCAGATATA ATTACAGAAA TAAATCCGTC TGTATCTTTT  
 1014 330 360  
 1015 TTACTGGAAG CTGATCCTAA GATAATATTA TCAAGGCAAA AGAGAGATAC AACAAGGAAT  
 1016 390 420  
 1017 AGAAATGATT ATAGTGACGA ATCAGTTATA TTAGAAACCA TAAACTTCGC TAGATATGCA  
 1018 450 480  
 30 1019 GCTACTGCTT CTGCAGTATT AGCCGGTTCT ACTGTTAAGG TAATTGTAAA CGTGGAAGGA  
 1020 510 540  
 1021 GATCCTAGTA TAGCAGCTAA TGAGATAATA AGGTCTATGA AGGGTGGTTC TTCTCTGGTT  
 1022 570 600  
 1023 CCGCGTGGAT CGGAATTCTG TTCATTTAGT TCTATCCCGA ATGGCACGTA CCGGGCGACG  
 1024 630 660  
 35 1025 TATCAGGATT TTGATGAGAA TGGTTGGAAG GACTTTCTCG AGGTTACTTT TGATGGTGGC  
 1026 690 720  
 1027 AAGATGGTGC AGGTGGTTTA CGATTATCAG CATAAAGAAG GGCGGTTTAA GTCCCAGGAC  
 1028 750 780  
 1029 GCTGACTACC ATCGGGTCAT GTATGCATCC TCGGGCATAG GTCCTGAAAA GGCCTTCAGA  
 40 1030 810 840  
 1031 GAGCTCGCCG ATGCTTTGCT TGAAAAGGGT AATCCCGAGA TGGTGGATGT GGTCACCGGT  
 1032 870 900  
 1033 GCAACTGTTT CTTCCCAGAG TTTCAGGAGG TTGGGTCGTG CGCTTCTGCA GAGTGC GCGG  
 1034 930 960  
 1035 CGCGGCGAGA AGGAAGCCAT TATTAGCAGG TAGGGATCCG GGCCCTCTAG ATGCGGCCGC  
 45 1036 990 1020  
 1037 ATGCATGGTA CCTAA  
 1038  
 1039  
 1040 (2) INFORMATION FOR SEQ ID NO: 21:  
 50 1041 (i) SEQUENCE CHARACTERISTICS:  
 1042 (A) LENGTH: 345 amino acids  
 1043 (B) TYPE: amino acid  
 1044 (D) TOPOLOGY: linear

55



1045 (ii) MOLECULE TYPE: protein  
 1046 (vi) ORIGINAL SOURCE:  
 1047 (A) ORGANISM: recombinant  
 5 1048 (x) PUBLICATION INFORMATION:  
 1049 (A) AUTHORS: Nobuyuki FUJII et al.  
 1050 (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID  
 1051 FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN  
 1052 (K) RELEVANT RESIDUES IN SEQ ID NO:21: FROM 1 to 345  
 1053  
 10 1054 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:  
 1055  
 1056 Met Lys Ile Gly Ile Val Thr Gly Ile Pro Gly Val Gly Lys Ser Thr Val Leu Ala Lys  
 1057 1 20  
 1058 Val Lys Glu Ile Leu Asp Asn Gln Gly Ile Asn Asn Lys Ile Ile Asn Tyr Gly Asp Phe  
 1059 21 40  
 15 1060 Met Leu Ala Thr Ala Leu Lys Leu Gly Tyr Ala Lys Asp Arg Asp Glu Met Arg Lys Leu  
 1061 41 60  
 1062 Ser Val Glu Lys Gln Lys Lys Leu Gln Ile Asp Ala Ala Lys Gly Ile Ala Glu Glu Ala  
 1063 61 80  
 1064 Arg Ala Gly Gly Glu Gly Tyr Leu Phe Ile Asp Thr His Ala Val Ile Arg Thr Pro Ser  
 1065 81 100  
 20 1066 Gly Tyr Leu Pro Gly Leu Pro Ser Asp Ile Ile Thr Glu Ile Asn Pro Ser Val Ile Phe  
 1067 101 120  
 1068 Leu Leu Glu Ala Asp Pro Lys Ile Ile Leu Ser Arg Gln Lys Arg Asp Thr Thr Arg Asn  
 1069 121 140  
 1070 Arg Asn Asp Tyr Ser Asp Glu Ser Val Ile Leu Glu Thr Ile Asn Phe Ala Arg Tyr Ala  
 1071 141 160  
 25 1072 Ala Thr Ala Ser Ala Val Leu Ala Gly Ser Thr Val Lys Val Ile Val Asn Val Glu Gly  
 1073 161 180  
 1074 Asp Pro Ser Ile Ala Ala Asn Glu Ile Ile Arg Ser Met Lys Gly Gly Ser Ser Leu Val  
 1075 181 200  
 1076 Pro Arg Gly Ser Glu Phe Cys Ser Phe Ser Ser Ile Pro Asn Gly Thr Tyr Arg Ala Thr  
 1077 201 220  
 30 1078 Tyr Gln Asp Phe Asp Glu Asn Gly Trp Lys Asp Phe Leu Glu Val Thr Phe Asp Gly Gly  
 1079 221 240  
 1080 Lys Met Val Gln Val Val Tyr Asp Tyr Gln His Lys Glu Gly Arg Phe Lys Ser Gln Asp  
 1081 241 260  
 1082 Ala Asp Tyr His Arg Val Met Tyr Ala Ser Ser Gly Ile Gly Pro Glu Lys Ala Phe Arg  
 1083 261 280  
 35 1084 Glu Leu Ala Asp Ala Leu Leu Glu Lys Gly Asn Pro Glu Met Val Asp Val Val Thr Gly  
 1085 281 300  
 1086 Ala Thr Val Ser Ser Gln Ser Phe Arg Arg Leu Gly Arg Ala Leu Leu Gln Ser Ala Arg  
 1087 301 320  
 1088 Arg Gly Glu Lys Glu Ala Ile Ile Ser Arg \*\*\* Gly Ser Gly Pro Ser Arg Cys Gly Arg  
 1089 321 340  
 40 1090 Met His Gly Thr \*\*\*  
 1091 341  
 1092

#### Claims

1. A fused DNA sequence which comprises a DNA sequence of a heat-resistant protein, fused directly or indirectly to a DNA sequence coding a selected protein or peptide.
2. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence derived from a thermophilic bacterium.
3. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence derived from a highly thermophilic bacterium.
4. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence derived from a *Thermophilus* bacterium, a *Sulfolobus* bacterium, a *Pyrococcus* bacterium, a *Thermotoga* bacte-

rium, a *Pyrobaculum* bacterium, a *Pyrodictium* bacterium, a *Thermococcus* bacterium, a *Thermodiscus* bacterium, a *Metanothermus* bacterium or a *Metanococcus* bacterium.

- 5
5. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence derived from a *Pyrococcus* bacterium or a *Sulfolobus* bacterium.
6. The sequence according to any one of Claims 1 to 5, wherein the DNA sequence of a heat-resistant protein is a DNA sequence of heat-resistant ferredoxin or heat-resistant adenyl kinase.
- 10
7. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence of ferredoxin derived from a *Pyrococcus* bacterium or adenyl kinase derived from a *Sulfolobus* bacterium.
8. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence of ferredoxin derived from a *Pyrococcus* bacterium.
- 15
9. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence of adenyl kinase derived from a *Sulfolobus* bacterium.
10. A fused protein which comprises being expressed from the DNA sequence according to any one of Claims 1 to 9.
- 20
11. A method for expressing a fused protein, which comprises using the DNA sequence according to any one of Claims 1 to 9.
- 25
- 30
- 35
- 40
- 45
- 50
- 55

Fig. 1

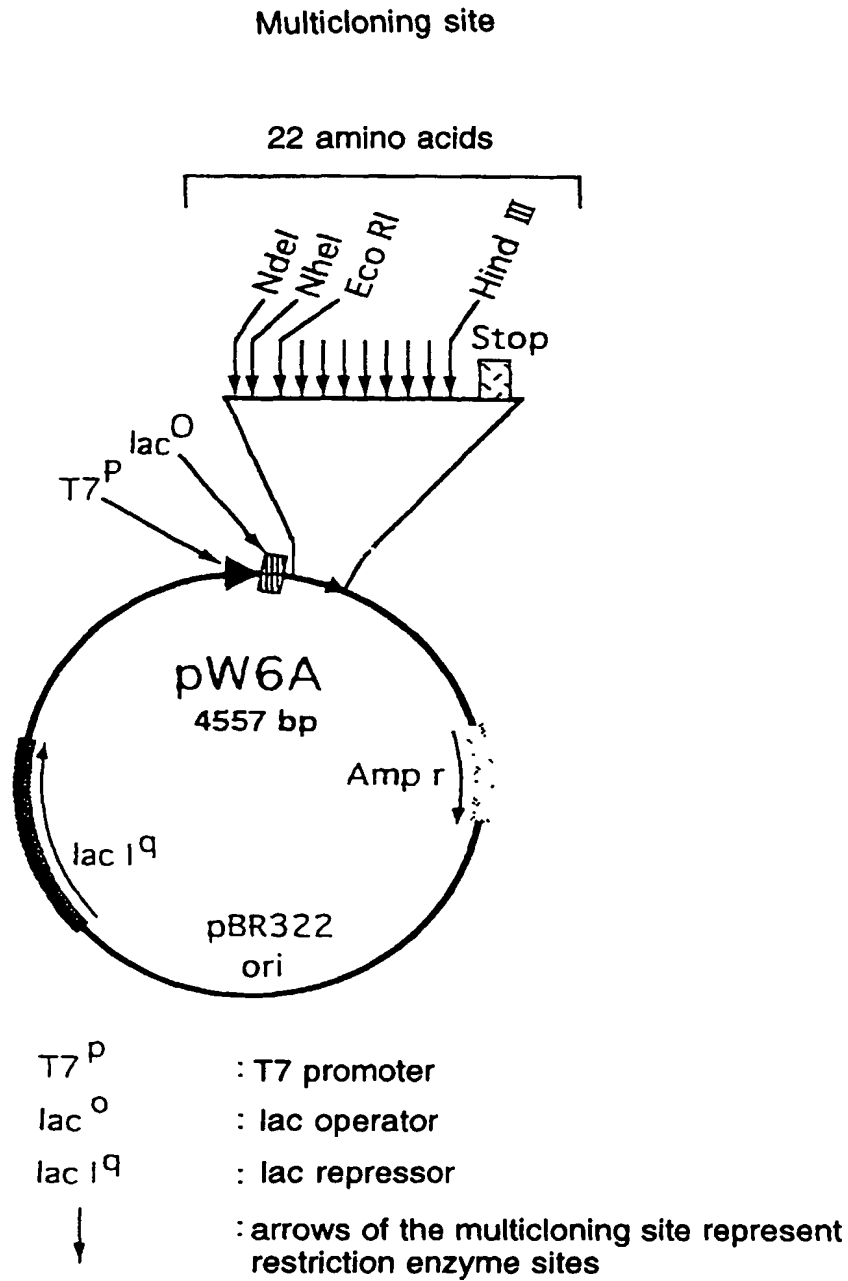
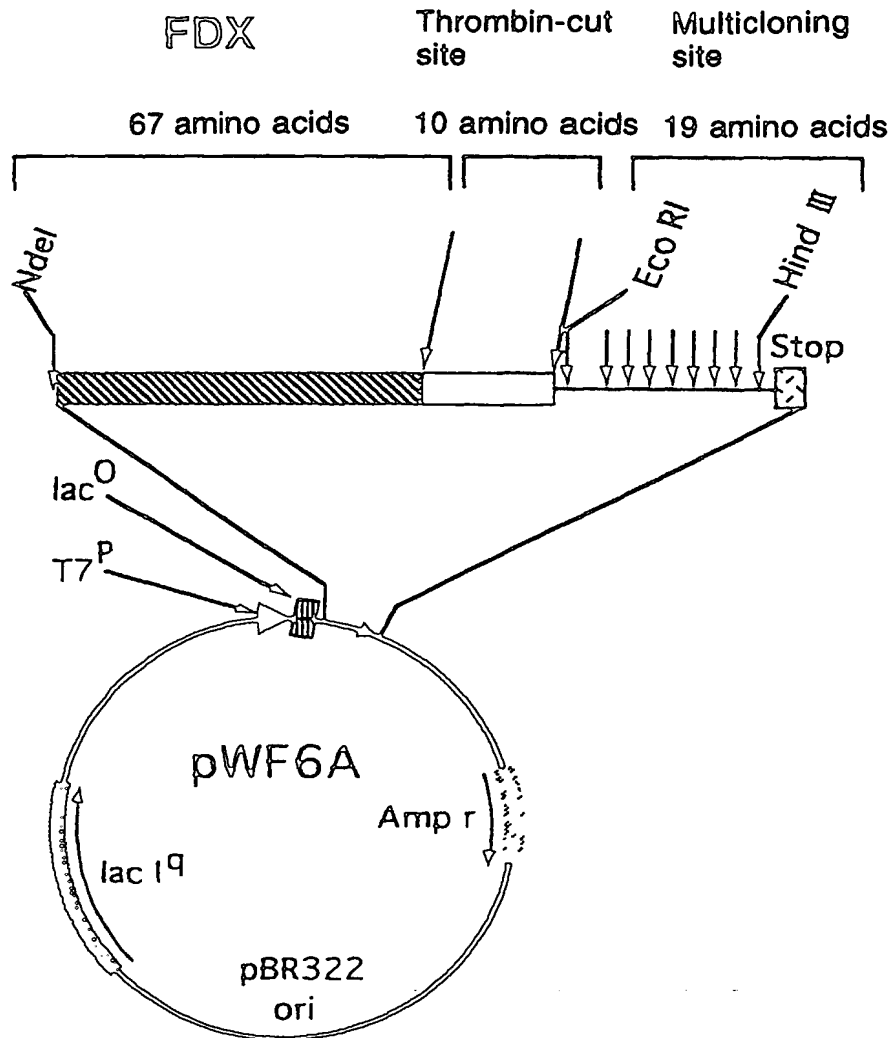


Fig. 2



- T7<sup>P</sup> : T7 promoter  
 lac<sup>O</sup> : lac operator  
 lac I<sup>q</sup> : lac repressor  
 ↓ : arrows of the multicloning site represent restriction enzyme sites

Fig. 3

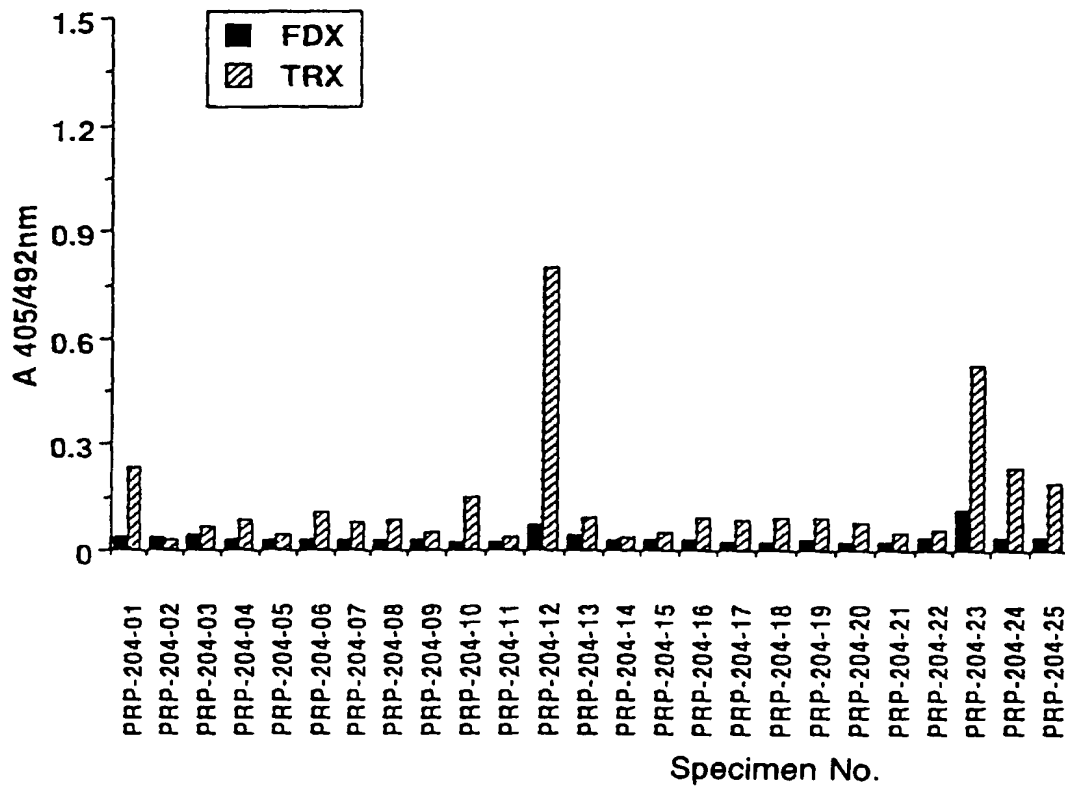


Fig. 4

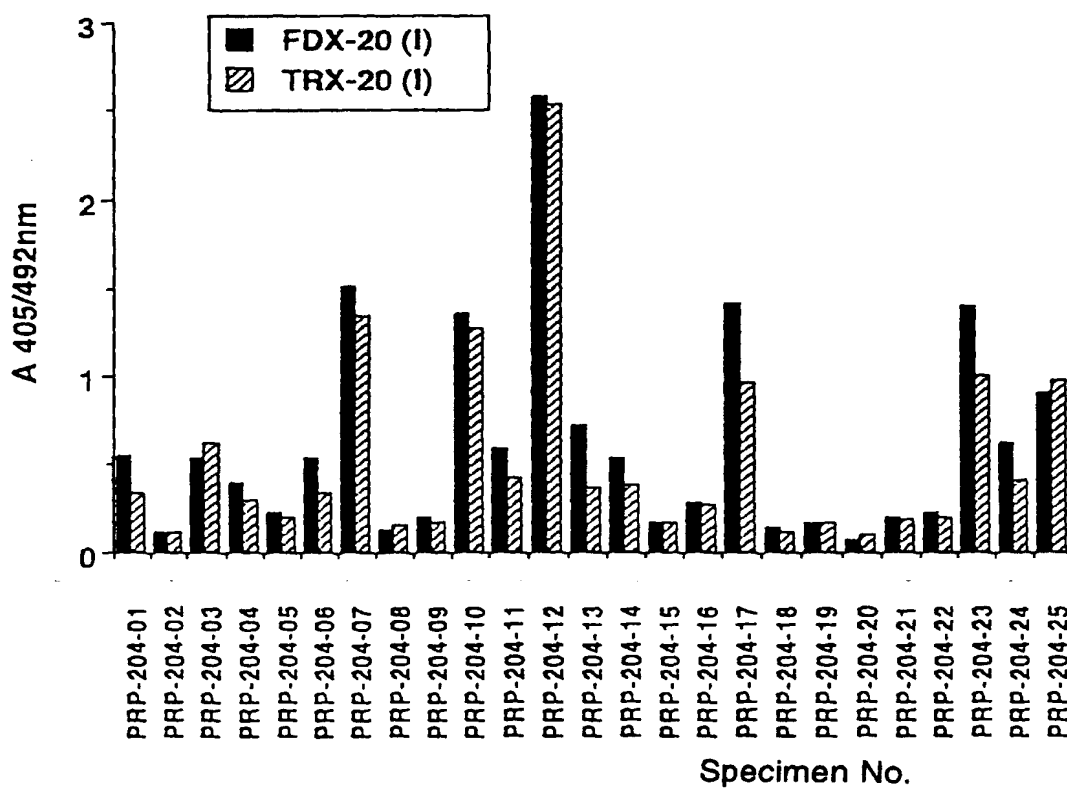


Fig. 5

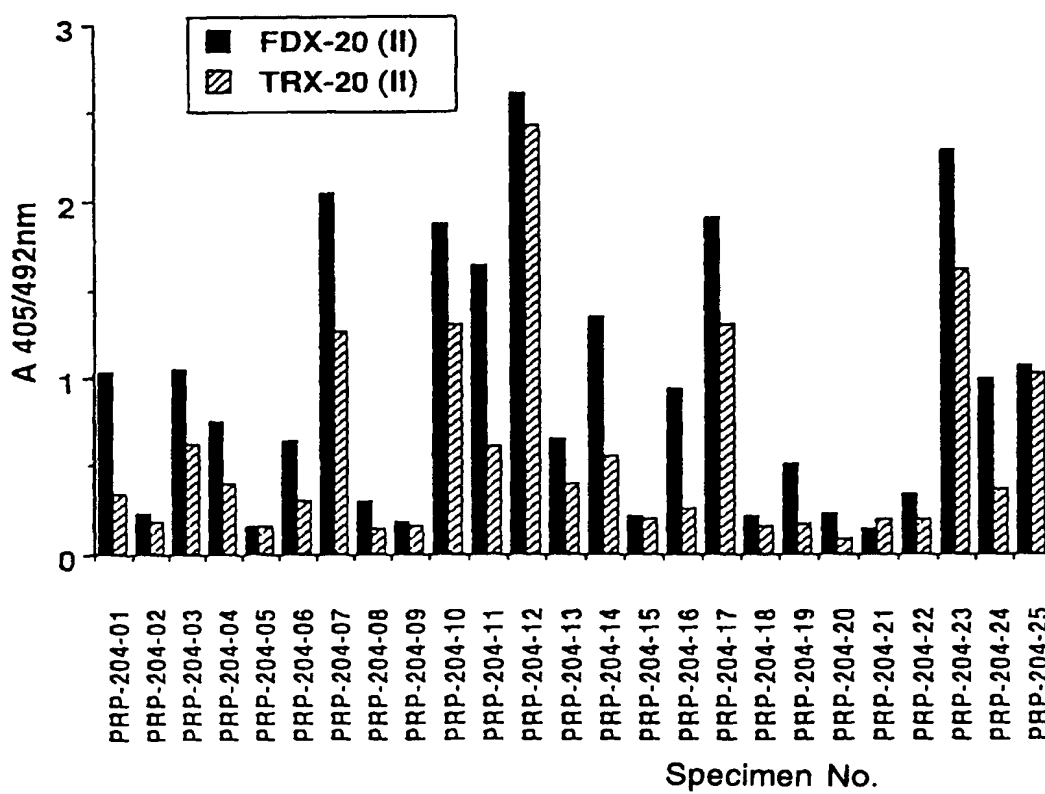


Fig. 6

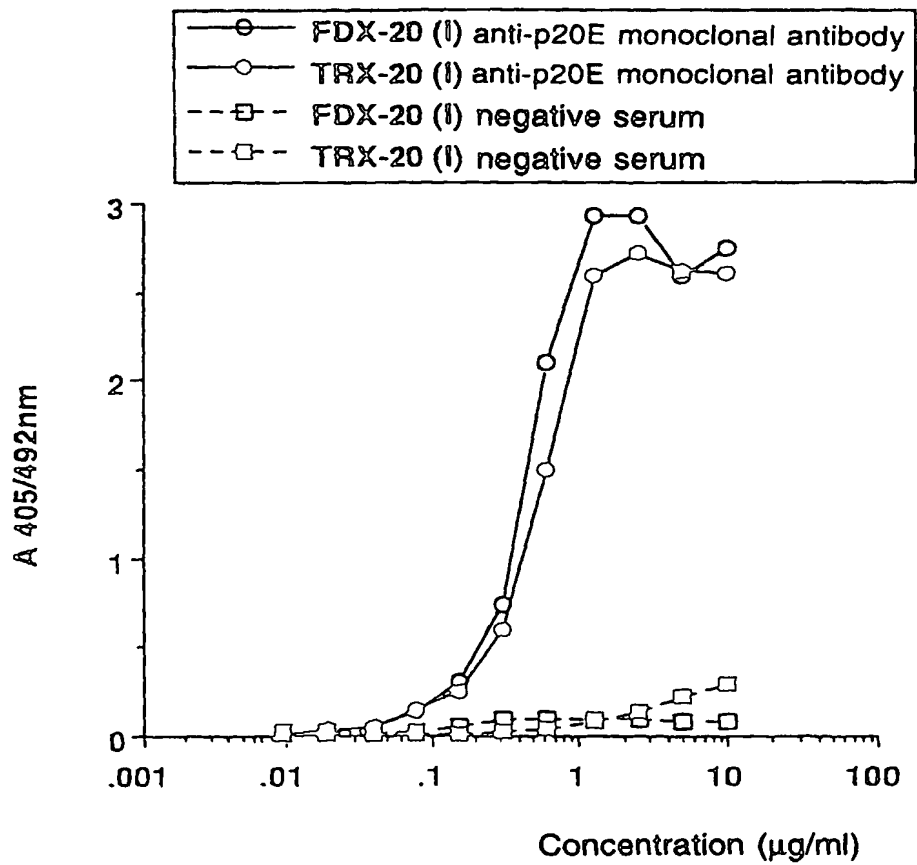




Fig. 7

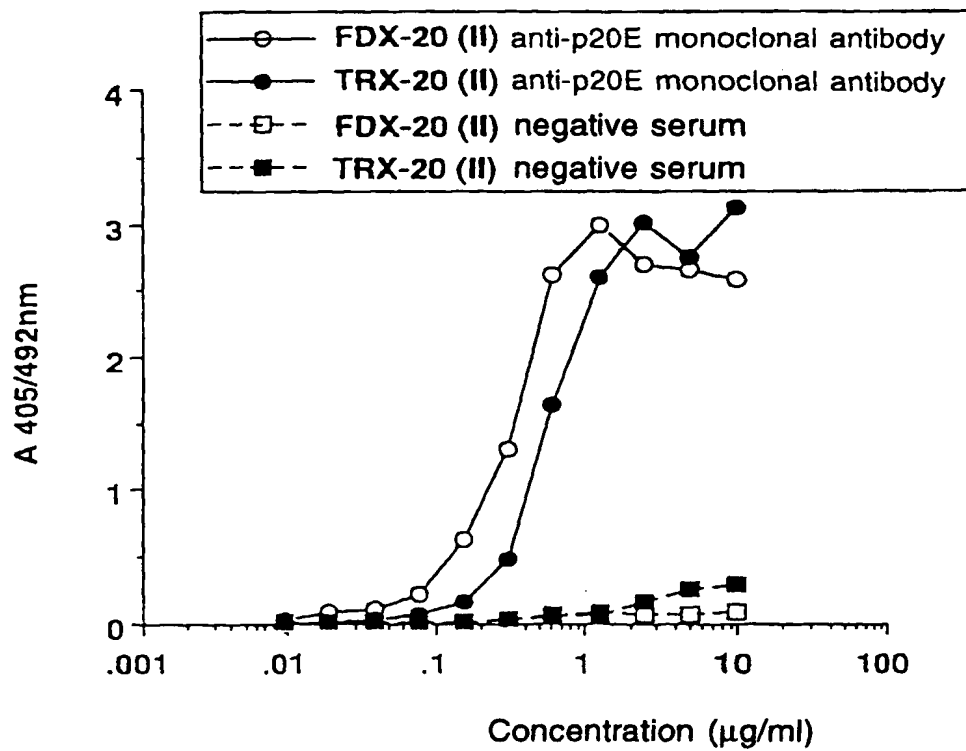


Fig. 8

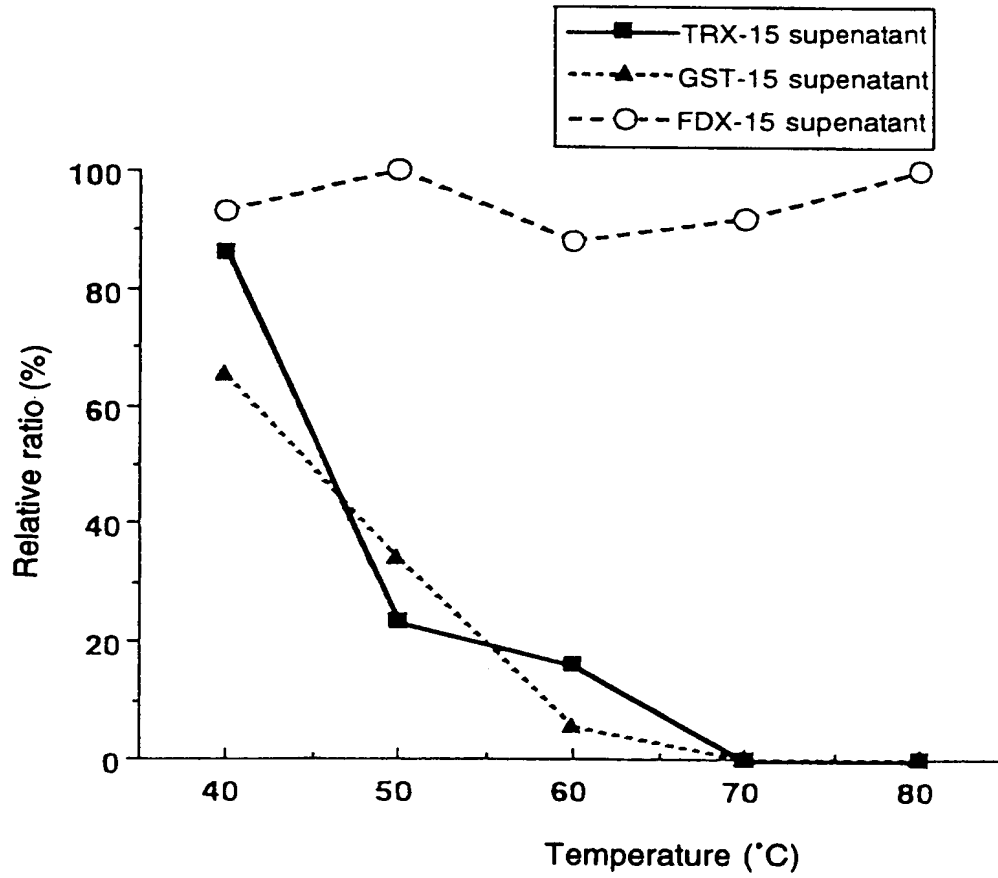


Fig. 9

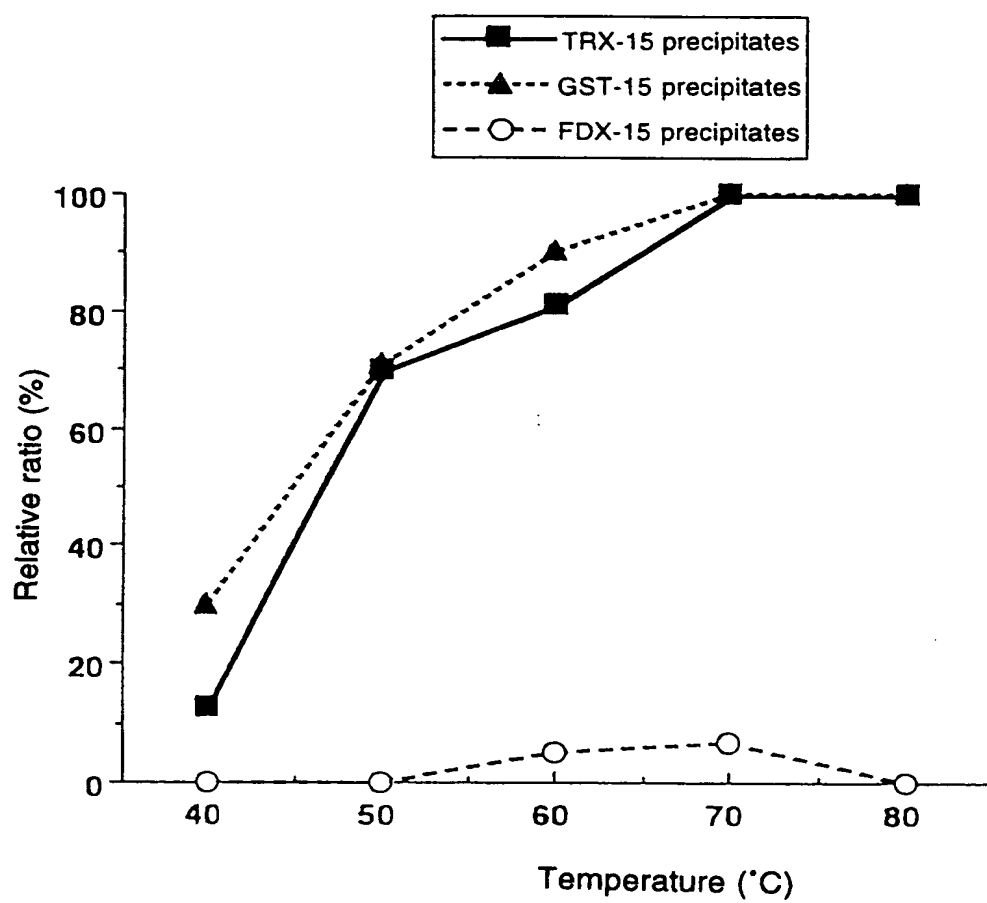
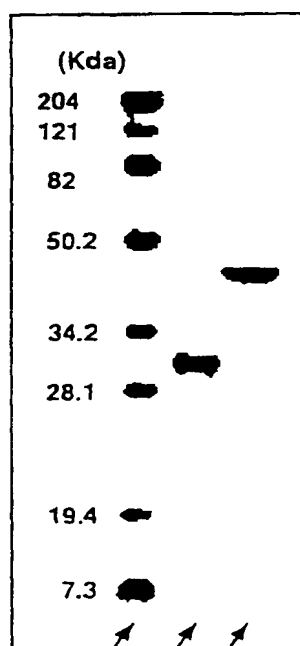


Fig. 10

CBB staining



Western blotting

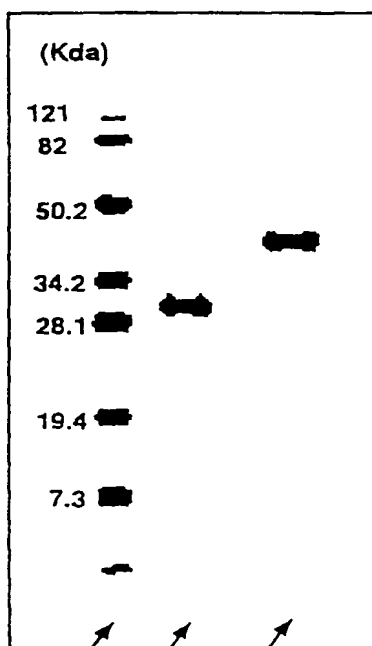


Fig. 11

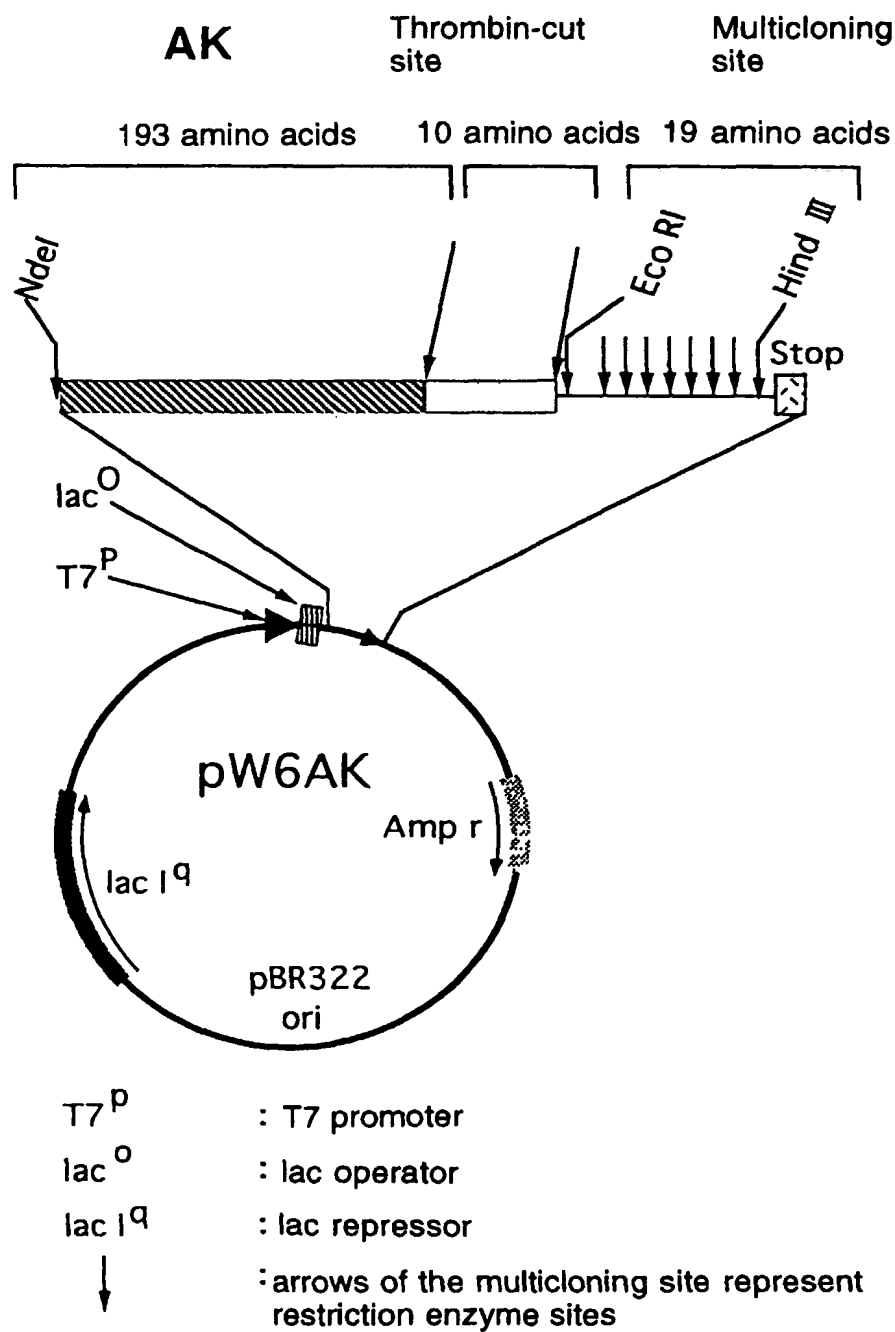
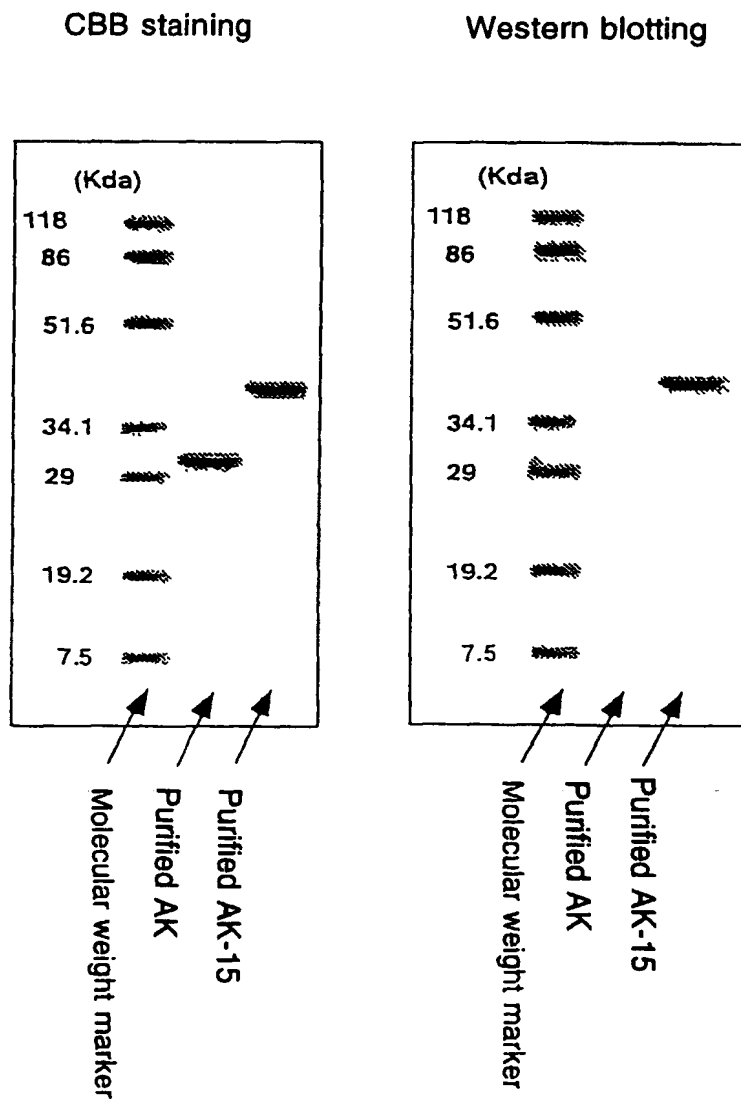


Fig. 12





Europäisches Patentamt  
Eur pean Patent Office  
Office européen des brevets



(11) EP 0 781 848 A3

(12) EUROPEAN PATENT APPLICATION

(88) Date of publication A3:  
01.10.1997 Bulletin 1997/40

(43) Date of publication A2:  
02.07.1997 Bulletin 1997/27

(21) Application number: 96120899.8

(22) Date of filing: 27.12.1996

(51) Int. Cl.<sup>6</sup>: C12N 15/62, C07K 14/195,  
C07K 14/15, C12N 9/12,  
C07K 14/20, C12N 15/31,  
C12N 15/48, C12N 15/54

(84) Designated Contracting States:  
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE

(30) Priority: 28.12.1995 JP 352225/95

(71) Applicant: FUJIREBIO Inc.  
Shinjuku-ku, Tokyo 163-07 (JP)

(72) Inventors:  
• Ueno, Eiichi,  
c/o Fujirebio Inc.  
Shinjuku-ku, Tokyo (JP)

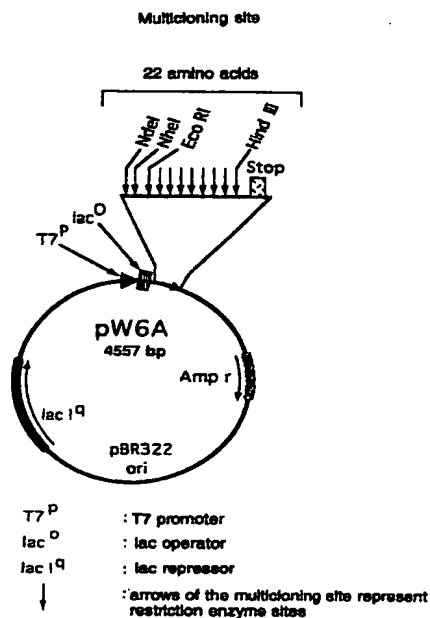
• Fujii, Nobuyuki,  
c/o Fujirebio Inc.  
Shinjuku-ku, Tokyo (JP)  
• Okada, Masahisa,  
c/o Fujirebio Inc.  
Shinjuku-ku, Tokyo (JP)

(74) Representative: Hansen, Bernd, Dr. Dipl.-Chem.  
et al  
Hoffmann Eitle,  
Patent- und Rechtsanwälte,  
Arabellastrasse 4  
81925 München (DE)

(54) Fused DNA sequence, fused protein expressed from said fused DNA sequence and method for expressing said fused protein

(57) Disclosed are a fused DNA sequence which comprises a DNA sequence of a heat-resistant protein, fused directly or indirectly to a DNA sequence coding a selected protein or peptide, a fused protein expressed from the fused DNA sequence, and a method for expressing the fused protein.

Fig. 1





European Patent  
Office

## EUROPEAN SEARCH REPORT

Application Number  
EP 96 12 0899

| DOCUMENTS CONSIDERED TO BE RELEVANT   |  |  |  |
|---|--|--|--|
| Category  | Citation of document with indication, where appropriate, of relevant passages  | Relevant to claim  | CLASSIFICATION OF THE APPLICATION (Int.Cl.6)   |
| X   | PROG. BIOTECHNOL. (1995), 10(CARBOHYDRATE BIOENGINEERING), 77-84 CODEN: PBITE3;ISSN: 0921-0423,<br>1995, XP000647006<br>MORACCI, M. ET AL: "Properties and production of the.beta.-glycosidase from the thermophilic Archaeon Sulfolobus solfataricus expressed in mesophilic hosts"   | 1-5,10, 11   | C12N15/62<br>C07K14/195<br>C07K14/15<br>C12N9/12<br>C07K14/20<br>C12N15/31<br>C12N15/48<br>C12N15/54 |
| Y   | * abstract *<br>* page 80, paragraph 4 - page 83, paragraph 3 *  | 6-9  |  |
| D,Y   | ---<br>ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS, vol. 307, no. 2, December 1993, pages 405-410, XP000647072<br>THOMAS KATH ET AL.: "Identification, cloning, and expression of the gene for Adenylate Kinase from the thermoacidophilic Archaeobacterium Sulfolobus acidocaldarius"<br>* abstract *<br>* page 406, right-hand column, paragraph 5 - page 408, left-hand column, paragraph 1 * | 6,7,9  |  |
| Y   | ---<br>J. BACTERIOL. (1994), 176(15), 4790-3<br>CODEN: JOBAAY;ISSN: 0021-9193,<br>1994, XP000670841<br>HELTZEL, ANDREAS ET AL: "Cloning, expression, and molecular characterization of the gene encoding an extremely thermostable [4Fe-4S] ferredoxin from the hyperthermophilic archaeon Pyrococcus furiosus"<br>* the whole document *  | 7,8  |  |
| The present search report has been drawn up for all claims  |  |  |  |
| Place of search<br><b>THE HAGUE</b>   |  | Date of completion of the search<br><b>7 August 1997</b>   | Examiner<br><b>Montero Lopez, B</b>  |
| <b>CATEGORY OF CITED DOCUMENTS</b><br>X : particularly relevant if taken alone<br>Y : particularly relevant if combined with another document of the same category<br>A : technological background<br>O : non-written disclosure<br>P : intermediate document |  | I : theory or principle underlying the invention<br>E : earlier patent document, but published on, or after the filing date<br>D : document cited in the application<br>L : document cited for other reasons<br>& : member of the same patent family, corresponding document |  |

EPO FORM 150 01.02 (P/E/C/D)





European Patent  
Office

# EUROPEAN SEARCH REPORT

Application Number  
EP 96 12 0899

| DOCUMENTS CONSIDERED TO BE RELEVANT  |   |  |  |
|--|---|--|--|
| Category   | Citation of document with indication, where appropriate, of relevant passages   | Relevant to claim  | CLASSIFICATION OF THE APPLICATION (Int.Cl.6) |
| A  | WO 87 05935 A (BATTELLE MEMORIAL INSTITUTE AT COLUMBUS) 8 October 1987<br>* page 16, line 6 - page 22, line 10 *<br>----- | 1,10,11  |  |
|  |   |  | TECHNICAL FIELDS SEARCHED (Int.Cl.6)         |
|  |   |  |  |
| The present search report has been drawn up for all claims   |   |  |  |
| Place of search<br><b>THE HAGUE</b>  |   | Date of completion of the search<br><b>7 August 1997</b> | Examiner<br><b>Montero Lopez, B</b>          |
| <p><b>CATEGORY OF CITED DOCUMENTS</b></p> <p>X : particularly relevant if taken alone<br/>Y : particularly relevant if combined with another document of the same category<br/>A : technological background<br/>P : non-written disclosure<br/>P : intermediate document</p> <p>T : theory or principle underlying the invention<br/>E : earlier patent document, but published on, or after the filing date<br/>D : document cited in the application<br/>L : document cited for other reasons<br/>-----<br/>&amp; : member of the same patent family, corresponding document</p> |   |  |  |

EPO FORM 150 (01.02) (P&C/01)

**THIS PAGE BLANK (USPTO)**